DEALING WITH UNCERTAINTIES IN INTEGRATED AGE-STRUCTURED ASSESSMENT MODELS

By

Peter-John Frederick Hulson

RECOMMENDED:

_____
Dr. Dana Hanselman

_____
Dr. Milo Adkison

_____
Dr. Gary Marty

_____
Dr. Brenda Norcross

_____
Dr. Terrance Quinn II, Advisory Committee Chair

_____
Dr. Milo Adkison
Chair, Graduate Program in Fisheries Division

APPROVED:

_____
Dr. Michael Castellini
Dean, School of Fisheries and Ocean Sciences

_____
Dr. Lawrence Duffy
Interim Dean of the Graduate School

_____
Date

DEALING WITH UNCERTAINTIES IN INTEGRATED AGE-STRUCTURED ASSESSMENT MODELS

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Peter-John Frederick Hulson, B.A., M.S.

Fairbanks, Alaska

May 2012

Abstract

Dealing with uncertainties in integrated age-structured assessment (ASA) models has become a central focus at all levels of fish stock assessment and management. My goal in this thesis was to uncover layers of uncertainty in ASA models. There were two major components to this approach: (1) dealing with uncertainty in datasets used in ASA models through examining effective sample size, and (2) dealing with uncertainty in ASA model structure through examination of effective number of parameters and simulations of school and stock structure. From the dataset uncertainty perspective, I investigated age and length composition datasets by first identifying possible sources of error and then by evaluating if it is feasible to include this error within ASA models. From the structural uncertainty perspective, I compared historical ASA models with spatially-explicit and metapopulation scale ASA models. Sources of uncertainty in age and length composition datasets include the spatial distribution of schools and age aggregation of fish within schools. To account partially for this error at the survey design level, the optimal approach is to sample from a greater number of schools, even if the sample size within any particular school is reduced. Also, it is possible to include this error within ASA models by parameterizing and estimating effective sample size with the Dirichlet distribution. Reduced uncertainty in parameters and management quantities resulted from spatially-explicit and metapopulation ASA models when compared to historical ASA model structures. Further, with possible climate change influences on fish populations use of spatially-explicit and metapopulation ASA models will allow stock assessment scientists to accurately and more precisely predict sustainable harvest levels.

Table of Contents

Page

Page

Page

Page

List of Figures

Page

Chapter 1

Chapter 2

Chapter 3

Page

Chapter 4

Page

Appendix A

List of Tables

Page

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Page

Appendix A

Acknowledgments

First and foremost I sincerely thank my major professor Dr. Terrance Quinn II. Sitting in Washington D.C. and talking about partial differential equations Terry generously took a chance on a young man who had the idea to apply math to fish populations. By taking that chance and offering me a graduate student position he changed the course of my life, for which I am deeply grateful. Throughout my time in graduate school he has always had a knack for challenging and rewarding me in just the right manner so that I neither became discouraged nor too full of myself. I look forward to enjoying many more years of conversations with Terry ranging from theoretical statistics to American Idol, and of course, helping fix his insufferable septic system. Terry is a coauthor on all my thesis chapters and for each he provided excellent advice, cajolement, and encouragement.

I also appreciate the skills, support, and guidance that each and every member of my committee provided. Drs. Gary Marty, Milo Adkison, Brenda Norcross and Dana Hanselman provided encouragement, friendship, and the tools to become a well rounded scientist. Dana was the second author on Chapters 1 and 2. A number of scientists in both the National Oceanographic and Atmospheric Administration and Alaska Department of Fish and Game were influential during my time in graduate school. This list includes Dr. Sherri Dressel (ADF&G) and Kyle Hebert (ADF&G) who were coauthors on Chapter 4. Some of the graduate students at UAF that struggled alongside me through our courses and provided mutual support since I first started include Katie Palof, Kray Van Kirk, Sara Miller, Cindy Tribuzio, William Bechtol, and Haixue Shen. I especially would like to

thank the office staff at SFOS, Debi Rathbone, Louisa Hayes, and Gabrielle Hazelton for always helping me whenever I needed it in a prompt and gracious manner.

I particularly would like to thank my family for their love and support. My mother, Dona, instilled a belief in me from my first breath that all things were possible and has proudly stood by me through both my achievements and failures. I thank my brother, Dan, and sister, Aireon, for always keeping me grounded; reminding me of who I am whenever my head starts to swell beyond my hat size. I especially thank my wife Megan for taking a chance with "dock boy" and agreeing to leave her home town to travel halfway across the globe to little rainy Juneau. My wife is my muse and none of this would have been possible without her inspiration, wit, charm, and grace. For providing me the initial opportunity to find the subject to which I would apply math I want to thank Ronald Vreeland. His leadership skills and work habits are traits I will always look up to, but I especially value our friendship. I also thank Ernie Panacek, for recognizing my abilities, developing them, and believing in me; he supported and encouraged me every step of the way.

This thesis is the result of research sponsored by Alaska Sea Grant with funds from the National Oceanic and Atmospheric Administration Office of Sea Grant, Department of Commerce, under grant no. NA06OAR4170013 (project number R/31-16) and grant no. NA10OAR4170097 (project number R/31-19), and from the University of Alaska with funds appropriated by the state. Funding was also provided by the Cooperative Institute for Arctic Research/Alaska Fisheries Science Center (CIFAR/AFSC) stock assessment fellowship program with funds from NOAA.

Dedication

I dedicate this thesis to the William Frederick Hulsons in my life: my father, Fred, and son, Liam. My father's passion for fishing rubbed off on me before I was out of diapers. In hindsight it seems fitting that I would become a fishery stock assessment scientist, even though neither of us had even heard of such an occupation just a handful of years ago. His life was an example of strength and endurance through painful and difficult times, without such an example I would not have been able to complete this dissertation. Even though I met my son just a handful of months ago he has been in the back of my mind throughout this process, and, he has helped me to understand my father more than I could have anticipated. I chose this path so that little Liam's challenges through life will be different than mine.

General Introduction

Analysis and evaluation of the status of fish stocks is one of the most difficult problems currently undertaken by fishery scientists. The primary reason for this difficulty lies in the nature of observation. In fisheries science, repeatable observations of population parameters are seldom possible, and highly unlikely events seem to occur quite often (Fournier *et al.* 1990). Due in part to the behavior of fish, the environment in which they live, and their relationship with unknown factors, observations can be logistically very difficult to obtain, and often irresolvable conflicts arise between data sources (Hulson *et al.* 2008). It can be argued that the most important activity in fisheries management is stock assessment: determining the status of a commercially or recreationally important fish population and what levels of harvest are sustainable. While the stock assessment process has evolved to develop models with increasingly complex biological processes (Quinn and Deriso 1999), there is still a need to improve and evaluate assessment models (NRC 1998, Quinn and Collie 2005).

A valuable tool available to fisheries scientists to understand and evaluate fish stocks are mathematical models. Models used for fisheries have developed from earlier studies on humans dating back to the 17th century (Caswell 2000, Quinn 2003). The initial development of modeling efforts in fisheries can be attributed to F. I. Baranov (Quinn 2003), who derived the theory of mortality for a year-class, which is separated into natural and fishing mortality. Early use of age-structured data in fisheries models was deterministic, and did not consider potential dataset error (Fry 1949, Jones 1964, Gulland 1965, Pope 1972). The first attempt to allow for uncertainty in fisheries age-

structured models was provided by Doubleday (1976), although the stochastic nature of the model was somewhat arbitrary. In 1982, a likelihood framework was constructed for age-structured assessment models (ASA, Fournier and Archibald 1982) allowing for methods to determine uncertainty in parameter estimates and population quantities to be developed and taken into account by fisheries managers. Deriso *et al.* (1985) developed a similar least squares approach and showed how the bootstrap method and retrospective analysis could be used to examine both data and model uncertainty.

The integrated ASA model has become the current gold standard for fish stock assessment (Quinn and Deriso 1999, chapter 8). Most major commercial fisheries around the world have ASA models as their centerpiece; just along the U.S. west coast, this includes Pacific halibut, most North Pacific and Pacific groundfish (pollock, cod, sablefish, rockfish, halibut, hake), and many Pacific herring stocks in the state of Alaska. The ASA model exists in many configurations around the world but they have several common features. The ASA model integrates multiple data sources and other relevant information including catch-at-age from fisheries and fishery-independent surveys, a survey index of abundance, and a fishery index of abundance. Observations are compared with model estimates in a statistical setting to obtain parameter estimates of recruitment, fishery and survey selectivity, fishing mortality, survey catchability, and sometimes natural mortality and maturity (e.g., Fu and Quinn 2000, Marty *et al.* 2003).

Dealing with uncertainties in integrated ASA models is multifaceted; within this thesis I focused on uncertainty in ASA models from two perspectives. From the first perspective, the dataset perspective, uncertainty in ASA models can be broadly assigned

into two categories: (1) uncertainty as it relates to observations to which ASA models are subsequently fit, and (2) uncertainty in model predictions that ultimately lead to recommendations on acceptable commercial or recreational catch limits that provide for sustainability of fish populations. From the second perspective, the structural perspective, uncertainty in ASA models can be interpreted as much more general, focusing more on the framework of assessment models and whether or not an ASA model pays sufficient attention to ecological effects of fisheries, for example. My goal in this dissertation was to investigate uncertainty and develop methods to deal with uncertainty in ASA models by evaluating both the uncertainty in datasets used in ASA models and uncertainty as it relates to the fundamental structure of the ASA model.

From the dataset perspective I investigated the potential sources of uncertainty in age and length observations that are used in ASA models. The dataset uncertainty and associated data weighting of age and length composition data is first a function of sample size. To accommodate unknown process or model misspecification error, many analysts specify a number for sample size that is smaller than the actual sample size, called the effective sample size (e.g., McAllister and Ianelli 1997, Hanselman *et al.* 2007). In chapter one, I investigated the sensitivity of age and length composition dataset uncertainty to potential sources of measurement, observation, process, and model specification error through investigation of effective sample size. Using the age and length composition datasets generated in chapter one, I evaluated several methods to parameterize effective sample size for composition data alone. I used the results from chapter one to estimate dataset uncertainty within ASA models for age and length

composition datasets in chapter two, which include multiple datasets for both composition data and abundance indices.

The structural perspective centered on evaluating ASA models across different spatial scales and developing a metapopulation ASA model that was based on oceanographic considerations. Climate change influences on the spatial distribution of fish stocks has become an increasing concern recently, and I tested the accuracy and precision of two spatial scales of ASA models that modeled a population whose spatial distribution was influenced by climate change (chapter three). I also investigated the feasibility of incorporating shared parameters for time-dependent deviations across regions in an ASA model to determine if metapopulation scale ASA models provide improvement over single stock ASA models in chapter four. Within the metapopulation model, movement among spatial regions was not considered like in the spatially explicit ASA model. Rather, a group of populations was integrated into the same ASA model so that broad scale influences on population dynamics could be investigated. In both chapters three and four the Deviance Information Criterion (DIC) was used for model comparison and the effective number of parameters was enumerated for comparison with the actual number of parameters. In chapter three the DIC value for spatially explicit ASA models was computed over a number of iterations to evaluate the range in potential DIC values when comparing model parameterizations. In chapter four DIC was used to compare between metapopulation ASA model parameterizations based on the geographic location of fish stocks. In both chapters, the effective number of parameters was calculated to deal with parameter identifiability.

In an effort to make the results of this dissertation more general I evaluated ASA models for three species: Pacific ocean perch (*Sebastes alutus*), walleye pollock (*Theragra chalcogramma*), and Pacific herring (*Clupea pallasii*). Pacific ocean perch are a long-lived rockfish species (maximum age of 84 observed in the Gulf of Alaska, Hanselman *et al.* 2003) that have a wide distribution in the North Pacific (Allen and Smith 1988). Walleye pollock are also broadly distributed throughout the North Pacific with the largest concentrations found in the Eastern Bering Sea (Ianelli *et al.* 2007) and represent over 40% of the global whitefish production. Pacific herring is a commercially and ecologically important forage fish species throughout most of its range along the shores of Asia and North America (Hay 1985).

Chapter 1: Effects of Process and Observation Errors on Effective Sample Size of Fishery and Survey Age and Length Composition Using Variance Ratio and Likelihood Methods[1]

Abstract

Observations of age or length composition from fisheries or research surveys are frequently modelled with the multinomial distribution. Violations of multinomial assumptions in data collection usually cause overdispersion of observations and consequent underestimation of uncertainty. This has led to the adoption of an effective sample size less than the actual sample size to approximate the likelihood function for age or length composition better in, for example, fishery stock assessment models. The behaviour of effective sample size is examined under different scenarios for population age distribution and sampling design. Effective sample size was approximated with three approaches: (i) the ratio of multinomial to empirical variance; (ii) sampling estimation; and (iii) the Dirichlet likelihood. The most significant changes in effective sample size were attributable to process error involving aggregation of ages within schools. In terms of observation error, effective sample size can be increased by increasing the number of tows from which samples are obtained for age or length composition, then, because of the reduced uncertainty in effective sample size, the Dirichlet likelihood can be integrated into the objective function of fishery stock assessment models to estimate effective sample size in future assessments.

[1] Hulson, P.-J. F., Hanselman, D. H., and Quinn, T. J., II, 2011. Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ratio and likelihood methods. ICES Journal of Marine Science 68(7): 1548-1557.

1.1 Introduction

Observations of age or length composition from fisheries or research surveys are frequently modelled with the multinomial distribution. Intuitively, the multinomial distribution describes the probability that a fish selected at random is a certain age or length class. For a dataset covering a number of years, the negative of the corresponding log-likelihood (omitting the constant combinatoric term) is (Quinn and Deriso 1999):

$$-\ln L = \sum_{Y} n_y \sum_{A} p_{a,y} \ln \hat{p}_{a,y} \,, \tag{1.1}$$

where n_y is the sample size in year y of Y years, $p_{a,y}$ the observed proportion of fish of age a from A ages obtained from sampling, and $\hat{p}_{a,y}$ is the predicted proportion obtained from a model. The principal assumption is that fish are sampled at random from the population, and that ages or lengths observed more often in the sampling are estimated with more precision. The multinomial variance of a proportion is inversely related to the sample size of fish taken for age or length composition, so, as sample size increases, the variance in the observed proportion decreases.

The effect of sample size on the variability of multinomial proportions taken from actual bottom trawl survey and commercial fishery observations is, however, more complicated than simple random sampling. Sampling methods and the behaviour of fish usually cause overdispersion of the true uncertainty in the estimated proportions (Coggins and Quinn 1998, Crone and Sampson 1998), so the uncertainty using actual sample size is an underestimation of the true error in the observations. In large part, this is because

fish tend to school with similar age groups; this positive intraclass correlation means that treating the sample as random will result in erroneously small uncertainty in the age or length distribution (Pennington and Vølstad 1994). Ultimately, there is no straightforward theoretical approach for deriving the true statistical distribution because of the lack of sufficient information to model these processes.

The use of the multinomial likelihood in fishery stock assessment models to fit observations of age or length composition distinguishes maximum likelihood methods (Fournier and Archibald 1982, Methot 1989) from least-square methods (Doubleday 1976, Deriso *et al.* 1985). In terms of model fitting, using the true sample size as the weighting term, n_y, in Equation (1.1) would force the model to fit observations of age or length composition too precisely, i.e. true sample sizes mask non-random sampling biases associated with such data-collection schemes. Overfitting age or length composition data can have a deleterious effect on the accuracy of related parameter estimates in fully integrated stock assessment models (recruitment, fishing mortality, selectivity, catchability, etc.), resulting in underestimated uncertainty in the estimated parameters and related management quantities (Patterson *et al.* 2001).

Because of this overdispersion, the weighting quantity that should be used in the multinomial likelihood, n_y, is smaller than the actual sample size of fish collected for age or length. This quantity is often referred to as the effective sample size (Pennington and Vølstad 1994, McAllister and Ianelli 1997). Like others, we define effective sample size as the number of samples taken for age or length composition that would be required for the variability based on simple random sampling to be the same as that based on more-

complex sampling designs and population characteristics (e.g. Pennington *et al.* 2002, Aanes and Pennington 2003). Studies have investigated survey or fishery sampling designs to improve the quality of observations, which for practical purposes increases the effective sample size (McGarvey and Pennington 2001, Cerviño and Saborido-Rey 2006).

In stock assessment models, proposals have been made either to estimate effective sample size from sampling variance (Pennington and Vølstad 1994, Crone and Sampson 1998) or to set effective sample size at a value lower than actual sample size (Methot 2000). Also, several methods have been proposed to fit age and length composition data that account for an effective sample size smaller than the observed sample size (McAllister and Ianelli 1997, Fournier *et al.* 1998).

Pacific Ocean perch (*Sebastes alutus*, POP) is a long-lived rockfish species found in the North Pacific and Bering Sea, with the largest populations around the Aleutian Islands, off northern British Columbia, and in the Gulf of Alaska (Allen and Smith 1988). The POP population is currently managed in the Gulf of Alaska with the use of a stock assessment model that fits observations of age and length composition with the multinomial likelihood (Hanselman *et al.* 2007). The two sources of age and length composition are from the commercial fishery and from the National Marine Fisheries Service (NMFS) bottom trawl survey. There may be significant overdispersion in the age or length composition, however, because POP exhibit seasonal differences in spatial and depth distributions (Love *et al.* 2002), being found in patchy, localized aggregations (Hanselman *et al.* 2001). Overdispersion of age or length observations may also be

attributed to aggregations of similar age within the schools sampled (McAllister and Ianelli 1997). Each of these processes would influence the effective sample size that should be used in the multinomial likelihood to fit the observations. Therefore, the Gulf of Alaska POP population provides an excellent candidate to evaluate the sensitivity of effective sample size for age and length composition.

Here, we investigate the behaviour of effective sample size with respect to process and observation errors in a simulated bottom trawl survey and commercial fishery. Using empirical data for age and length composition from the Gulf of Alaska POP bottom trawl survey and commercial fishery, a mixture of parametric and non-parametric resampling methods (Manly 1997) is applied. The key considerations for process error in the simulations included the spatial distribution of fish age associated with habitat, spatial clustering of fish within habitats, and age aggregation within schools. Considerations of observation error included spatial placement of tows, ageing error, and subsampling of the catch. Our objectives were twofold: (i) to evaluate effective sample size over replicates of simulated age and length composition with the proposed methods, and (ii) to investigate the relationship between effective sample size and sources of process and observation error.

1.2 Materials and methods

1.2.1 Approximation approaches to determining effective sample size

Three methods to approximate effective sample size ($\breve{n}_y$) of the bottom trawl survey and commercial fishery were investigated. The methods used were selected to

represent two broad categories: (i) methods that relate different sources of variability (variance ratio methods), and (ii) maximum likelihood. In each approximation method, the known "true" proportions for the survey or fishery age composition and fishery length composition ($p_{a,y}$ or $p_{l,y}$) were taken from the 2007 age-structured assessment (ASA) model for POP. The estimated proportions ($\hat{p}_{a,y}$ or $\hat{p}_{l,y}$) were then obtained from simulation scenarios that included different sampling designs and population characteristics. Effective sample size was determined using the simulated data for each of the three methods, and the expected value of effective sample size from each approximation method was the average over the number of replicate surveys or fisheries. The details of the three methods are: (i) for the first variance ratio method, effective sample size was calculated directly; (ii) for the second variance ratio method, data from the simulated sampling for age/length composition were used to calculate effective sample size; and (iii) for the maximum likelihood method, effective sample size was estimated using code that linked optimization in AD Model Builder (ADMB, Fournier 1996) to data-generation functions written in R (R Development Core Team 2008).

The first variance ratio method (multinomial approximator) to determine effective sample size was based on making the ratio of the multinomial variance in the estimated proportions equal to an empirical approximation of variance between the observed and estimated proportions. Solving for sample size, the multinomial approximator of effective sample size, ${}_{M}\breve{n}_{y}$, derived in McAllister and Ianelli (1997) is

$$ {}_{M}\breve{n}_{y} = \frac{\sum_{a=2}^{A} \hat{p}_{a,y}\left(1-\hat{p}_{a,y}\right)}{\sum_{a=2}^{A}\left(p_{a,y}-\hat{p}_{a,y}\right)^{2}}, \tag{1.2} $$

where $\hat{p}_{a,y}$ is the estimated proportion at age *a* in year *y*, and $p_{a,y}$ is the known "true" proportion (ages 2–25). To calculate effective sample size for length composition data, the subscripts for age *a* were simply replaced by length *l*.

The second variance ratio method (sampling approximator) defines effective sample size so that the sampling variability based on a simple random sample would be the same as the variability based on the cluster samples (Pennington *et al.* 2002). This method was selected because in the first variance ratio method and in the maximum likelihood method, the known "true" age/length composition of the population must be used to approximate effective sample size. Using this method allowed for approximation of effective sample size when the "true" age/length composition is unknown. Following slightly different notation from Pennington *et al.* (2002), we calculated a ratio estimator of the mean age given by

$$ \hat{R}_{y} = \frac{\sum_{t=1}^{T} C_{t,y}\hat{\mu}_{t,y}}{C_{y}}, \tag{1.3} $$

where $C_{t,y}$ is the number of fish caught in tow t for year y, $\hat{\mu}_{t,y}$ an estimate of the average age or length of the sample in tow t, and C_y is the total catch in year y (i.e. the sum of the catch from each tow, $C_{t,y}$). The estimated variance of the ratio estimator is approximately

$$\mathrm{var}\left(\hat{R}_y\right) = \frac{\sum_{t=1}^{T}\left(C_{t,y}/n_{t,y}\right)^2\left(\hat{\mu}_{t,y} - \hat{R}_y\right)^2}{T(1-T)}, \tag{1.4}$$

where $n_{t,y}$ is the number of fish sampled in tow t, and T is the total number of tows. The variance of the age or length distribution if the samples are selected randomly is estimated as

$$\hat{\sigma}_y^2 = \frac{\sum_{t=1}^{T}\sum_{j=1}^{n_{t,y}}\left(C_{t,y}/n_{t,y}\right)\left(x_{t,j,y} - \hat{R}_y\right)^2}{C_y - 1}, \tag{1.5}$$

where $x_{t,j,y}$ is the age of the jth fish in tow t. The sampling approximator of effective sample size, ${}_S\breve{n}_y$, is then (Pennington *et al.* 2002)

$${}_S\breve{n}_y = \frac{\hat{\sigma}_y^2}{\mathrm{var}\left(\hat{R}_y\right)}. \tag{1.6}$$

The maximum likelihood method (Dirichlet approximator) utilizes the Dirichlet distribution and estimates effective sample size based on maximum likelihood techniques. In previous research, the Dirichlet likelihood was used to approximate effective sample size to evaluate uncertainty in modelled quantities (Williams and Quinn 1998), and as a component of the likelihood function in stock assessment models (Schnute and Haigh 2007). In this study, these methods were used to evaluate the effective sample size of the simulated bottom trawl survey and commercial fishery. The Dirichlet is the conjugate prior to the proportions in the multinomial likelihood (Gelman *et al.* 2004) and a continuous distribution for proportions, so represents a natural likelihood to evaluate effective sample size. In this approach, the Dirichlet parameter was equivalent to the effective sample size, and the Dirichlet approximator of estimated effective sample size, ${}_D\breve{n}_y$, was obtained by minimizing the negative log-likelihood:

$$-\ln L_y = -\left[\ln\Gamma\left({}_D\breve{n}_y\right) - \sum_{a=2}^{A}\ln\Gamma\left({}_D\breve{n}_{a,y}\right) + \sum_{a=2}^{A}\left({}_D\breve{n}_{a,y}-1\right)\ln p_{a,y}\right], \qquad (1.7)$$

where ${}_D\breve{n}_{a,y} = \hat{p}_{a,y}\times{}_D\breve{n}_y$ for $a = 2,\ldots,A–1$, and for the last age group ${}_D\breve{n}_{A,y} = {}_D\breve{n}_y - \sum_{a=1}^{A-1}{}_D\breve{n}_{a,y}$.

1.2.2 Simulation scenarios

The simulations to evaluate effective sample size were generated from data obtained for POP in the NMFS bottom trawl survey and commercial fishery. Tow sample

data, including depth of tow, number of tows with catch, total catch, number of tows from which fish were sampled, and sample size, were compiled for the bottom trawl survey and commercial fishery for years in which data are available (bottom trawl survey age: 1984, 1987, 1990, 1993, 1996, 1999, 2003, 2005, commercial fishery age: 1998–2002, 2004–2006, commercial fishery length: 1991–1992, 1995–1997). The known "true" population composition upon which the simulated bottom trawl survey and commercial fishery samples were based was taken from the 2007 stock assessment for Gulf of Alaska POP (Hanselman *et al.* 2007).

In total, 100 simulation scenarios evaluated the introduction of process and observation errors to sampling for ages and lengths from the bottom trawl survey and commercial fishery. In this study, the main attributes of catching and sampling fish in the bottom trawl survey or commercial fishery were simulated. Overall, the simulation was a mixture of parametric and non-parametric resampling methods that have been used previously to test fishery surveys (Sigler and Fujioka 1988, Schnute and Haigh 2003).

Sampling for age or length composition from the population through either the bottom trawl survey or commercial fishery consisted of four primary steps selected to imitate the actual process by which fish are sampled. This data-generation process determined (1) the depth stratum of the tow, (2) whether the tow contained fish, (3) how many fish were sampled from the tow for length and age determination, and (4) the age or length classes of the sample. These steps with associated scenarios are illustrated in Figure 1.1. Nine observation error and five process error scenarios were investigated (Table 1.1). Cases denoted with the subscript 0 (i.e. D_0) within the process error and

observation error scenarios represented the cases most similar to current methodologies, and cases with subscripts ≥1 represented alternative methods or sources of error.

Two scenarios of observation error in step 1 considered the spatial placement of tows with respect to depth strata (Table 1.1), because the spatial distribution of the ages of POP may be associated with depth (Love *et al.* 2002). These cases were selected to determine the influence of potentially undersampling certain depths on estimating effective sample size. In case D_0, depth strata towed in each year by the bottom trawl survey and commercial fishery were resampled with replacement. In case D_1, all depth strata were selected with equal probability (i.e. uniform placement of tows within depth strata).

In step 2, two scenarios of process error evaluated the influence of spatial clustering of fish. There are a number of tows in the bottom trawl survey that do not catch fish because of a combination of random placement of tows and clustering of fish. Alternatively, commercial fishery operations utilize electronics to identify fish schools prior to setting out tow gear, so nearly every tow in the fishery catches fish. By comparing the bottom trawl survey and commercial fishery, the impact of spatial clustering on resulting estimates of effective sample size was evaluated. To imitate spatial clustering of fish, the binomial distribution was used to determine whether the tow was successful, i.e. whether there were fish present in the haul. A binomial random variable was generated based on the proportion of tows with fish, calculated from the bottom trawl survey and commercial fishery data. If the tow did not contain fish, the simulation started again for the next tow. If the tow was determined to be successful, step

3 was performed to determine if a sample of fish was to be taken for ages or lengths. This process is representative of the delta-lognormal method commonly used to model survey and fishery catch per unit effort (cpue) where there are many zero catches (Lo *et al.* 1992).

Step 3 included five observation error scenarios that investigated the effect of the number of tows from which fish were sampled and the sample size for age or length compositions. In step 3, similar to step 2, a binomial random variable was generated to determine whether a sample of fish was to be taken for age or length composition from a successful tow. The binomial draw was based on the proportion of successful tows from which samples of fish were taken for age or length composition in the bottom trawl survey and commercial fishery data. If a sample was not taken, the simulation continued to the next tow. If a sample was taken for subsequent age and length composition, the total sample size was subjected to sampling with replacement from observer data in the bottom trawl survey and commercial fishery. The five cases in step 3 were determined by multiplying the proportion of tows from which fish were sampled in combination with the number of fish sampled, by 0.5, 1, or 2. The first, S_0, was selected to generate total sample sizes similar to the actual sample sizes obtained. The second and third cases, S_1 and S_2, resulted in total sample sizes similar to actual sample size, but based on the samples distributed across tows differently. In cases S_3 and S_4, the total sample size was increased by either doubling the sample size in each tow (case S_3), or doubling the number of tows sampled (S_4).

Two scenarios for observation error in step 4 investigated the influence of ageing error, which has an effect on the precision of age compositions derived from sampling (Coggins and Quinn 1998). The first, E_0, allowed for ageing error in the samples using the ageing error applied in the 2007 stock assessment to the bottom trawl survey and commercial fishery age data (Hanselman *et al.* 2007). In the second, E_1, ageing error was not present in the samples (Table 1.1).

In all, five scenarios of process error (Table 1.1) examined the influence of aggregations of ages within schools and school spatial distribution with respect to depth strata in step 4. Age aggregation within schools was considered in three cases, in which schools were composed of one age, mixed ages (standard deviation of ages within schools set at 10), and just one set of mixed ages comprising the known population age composition. Two cases investigated the spatial distribution of schools where schools were distributed randomly across depth strata (A_1, A_3), and two cases investigated mean age in the school increasing with depth strata (A_2, A_4). The spatial distribution of schools was not considered when the schools consisted of the known age composition (A_5), because in this case, the composition of a school was the same in all depth strata, so the estimated effective sample size was not affected.

For each age aggregation case in step 4, the proportion-at-age/length of fish within depth strata was based on a mean and standard deviation of ages within the school, i.e. an age-aggregation matrix $\Phi_{a,d}$. The age-aggregation matrix was a 24 × 24 matrix, with dimensions that included age and depth strata. For example, the age-aggregation matrix from case A_2 was simply the identity matrix (matrix of ones on the diagonal and

zeros elsewhere), and from A_5 was a matrix with entries all equal to 1. For each depth stratum, the "true" population composition obtained from the 2007 stock assessment was combined with the age-aggregation matrix to provide the composition of ages and lengths within schools, by depth, for each year in the survey and fishery ($\Omega_{a,y,d}$). This was calculated as

$$\Omega_{a,y,d} = \frac{\Phi_{a,d}\, p_{a,y}}{\sum_{A} \Phi_{a,d}\, p_{a,y}}, \tag{1.8}$$

where the proportion-at-age in the school, given by the age-aggregation matrix, $\Phi_{a,d}$, was multiplied by the proportion-at-age within the population, $p_{a,y}$, then scaled so that the proportions-at-age within a school in depth stratum *d* summed to 1. The number of fish sampled from each tow was then used as the sample size to generate school proportion-at-age/length in depth strata (based on a multinomial error distribution), to obtain samples by age or length.

In each replicate of the bottom trawl survey and commercial fishery, the number of fish-at-age/length was compiled for the tows, resulting in an estimated age ($\hat{p}_{a,y}$) or length ($\hat{p}_{l,y}$) composition. The number of tows was set to the actual number of tows for each year in the bottom trawl survey and commercial fishery. Given the simulated bottom trawl survey and commercial fishery age and length observations, and the true proportions based on the POP 2007 stock assessment model, the effective sample size for

the replicate was then determined from the three estimation approaches. The number of replicates for generating age or length compositions from the annual bottom trawl survey and commercial fishery was 1000, which allowed for accurate determination of the expected effective sample size (hereafter referred to as effective sample size for brevity) by each of three approximation methods (Efron and Tibshirani 1993, Aanes and Pennington 2003).

1.3 Results

In general, the trends in effective sample size across process error scenarios were similar between the multinomial approximation and the Dirichlet approximation methods. Investigating the case most comparable with the actual sampling of the bottom trawl survey (case $D_0 - E_0 - S_0$) for process error scenarios $A_{1,3,5}$ resulted in an increase in effective sample size because schools contained increasing numbers of ages (Figure 1.2). Further, the uncertainty in effective sample size from the sampling approximator was generally larger than for the multinomial or Dirichlet approximators. For all simulation scenario cases of the bottom trawl survey and commercial fishery, the multinomial approximator produced nearly identical effective sample sizes to the Dirichlet approximator (top panel, Figure 1.3). The general trend in the standard deviation of effective sample size indicated that lower variance can be obtained with the Dirichlet approximator than with the multinomial approximator (bottom panel, Figure 1.3). For the remaining results, the Dirichlet approximator was selected as the preferred method, mainly to eliminate redundancy, but also because of the reduced variability in effective

sample size determinations. For comparison among process error and observation error scenarios, most of the results are shown for cases that are the most similar to current methods, unless directly compared with alternative cases. Results for ageing error and sampling of catch cases are shown as ratios using the log (base 2) function (e.g. Schnute and Richards 1995). A ratio with a value of 1 in ageing error results, for example, is interpreted as the effective sample size in case E_0 being twice as large as the effective sample size in case E_1, a value of 3 would 8 times as large, and so on.

In most years of the bottom trawl survey, effective sample size increased from process error scenario A_1 to A_5 as a consequence of the increasing numbers of ages in schools (Table 1.2). The smallest effective sample size was in process error scenario A_1, in which the schools consisted of a single age and were distributed randomly with respect to depth strata. Alternatively, the largest effective sample size resulted in values that were nearly equal to the total sample size in process error scenario A_5, where all schools consisted of the true age composition. Effective sample size for process error scenario A_2, in which schools consisted of one age and the mean school age increased with depth strata, were larger than for A_1 and were comparable in magnitude for most years for mixed-age process error scenarios A_3 (random distribution of schools) and A_4 (increasing mean age of school with depth).

Effective sample size in the commercial fishery age and length composition also increased with increasing age aggregation of schools (Table 1.2). Similar to the bottom trawl survey, the smallest effective sample size for the commercial fishery age composition data resulted from process error scenario A_1 and the largest from A_5. Process

error scenario A_2 resulted in nearly the same effective sample size in all years as A_1. Further, in all years the effective sample size from process error scenario A_3, in which schools are distributed randomly with depth strata, was an order of magnitude larger than that from process error scenario A_4, in which the mean age increases with depth strata (Table 1.2). For the commercial fishery length composition, effective sample size was mostly larger when schools were distributed randomly across depth strata than when ages increased with depth, regardless of the age aggregation within the school (Table 1.2).

In general, the effective sample size for the bottom trawl survey and commercial fishery compared among process error scenarios A_1–A_4 were within an order of magnitude (Table 1.2). Comparing between age composition from the bottom trawl survey and commercial fishery, effective sample size for most years from process error scenario A_3 was larger in the commercial fishery, and effective sample size from A_4 was larger in the bottom trawl survey. Overall, the magnitude of effective sample size from the bottom trawl survey age composition was similar to the commercial fishery age composition. For most years, effective sample size for length composition from the commercial fishery was similar to that for age composition from the bottom trawl survey and commercial fishery, even though the actual sample size for lengths was nearly double that of the bottom trawl survey and 3–5 times larger than the commercial fishery age composition. There were exceptions, however. The effective sample size for the commercial fishery length composition was an order of magnitude larger in 1996 and 1997 for process error scenario A_3 than effective sample size for the bottom trawl survey or commercial fishery age composition in any year for process error scenarios A_1–A_4

(Table 1.2). The commercial fishery length composition effective sample size was also larger in process error scenario A_4 for each year compared with the same process error scenario for the commercial fishery age composition.

Effective sample size from the bottom trawl survey in the observation error scenarios was sensitive to spatial distribution of POP schools with regard to the pattern in sampling depth. Effective sample size in case D_0 was much higher for process error scenarios A_3–A_4 than for A_1–A_2, with smaller changes within pairs and different directional changes (left panel, Figure 1.4). Effective sample size in case D_1 was also much higher for process error scenarios A_3–A_4 than for A_1–A_2, but there were no changes within pairs (right panel, Figure 1.4). Effective sample size was often different among years too.

When ageing error was introduced, effective sample size in both the bottom trawl survey and commercial fishery was increased when schools consisted of a single age (scenarios A_1 and A_2, Figure 1.5). In process error scenarios A_1 and A_2, the ratios resulted in positive values and effective sample size that was up to four times larger in case E_0 than in E_1. This result is counter-intuitive, but involves the mixing of process error and observation error when sampling from schools consisting of just a single age class. Effective sample size was similar in ageing error cases E_0 and E_1, with most ratios near zero for schools of mixed age (scenarios A_3 and A_4). However, the mean ratio was slightly positive, indicating a small increase in effective sample size when ageing error was present (Figure 1.5). In the bottom trawl survey and to a lesser extent the commercial

fishery, the effective sample size in case E_0 was slightly smaller than in case E_1 (negative ratio) in scenario A_5.

In general, effective sample size from the bottom trawl survey and commercial fishery, for age and length composition, increased as the number of tows sampled increased (Figure 1.6). Ratios in effective sample size among the observation error scenarios for sampling of catch cases S_1–S_4 were considered for schools consisting of a single age (process error scenarios A_1 and A_2, left panel, Figure 1.6) and mixed ages (A_3 and A_4, right panel, Figure 1.6) and were calculated with the $\log_2$ function for comparison with base case S_0. For schools consisting of a single age, effective sample size decreased for bottom trawl survey and commercial fishery age composition when fewer tows were sampled and more fish were taken from each tow, case S_1, but did not change in the commercial fishery length composition compared with base case S_0 (left panel, Figure 1.6). For schools of mixed age, there was an increase in effective sample size for each dataset in case S_1 (right panel, Figure 1.6). Decreasing the number of fish sampled in each tow while increasing the number of tows sampled in case S_2 resulted in an increase in effective sample size for all datasets considered for schools composed of a single age, but for schools of mixed age, the effective sample size increased only for the bottom trawl survey age composition. As expected, increasing the overall total sample size by doubling the amount of fish sampled in each tow (S_3) resulted in larger effective sample sizes than the base case for both the bottom trawl survey and commercial fishery for schools consisting of single and mixed ages (Figure 1.6). Similar results were obtained when the same number of fish were sampled in each tow of the observer data, but twice

the number of tows were sampled (S_4), although in this case, the effective sample size increased from that in case S_3 when schools consisted of a single age (Figure 1.6).

1.4 Discussion

Overall, the aggregation of ages within fish schools and the spatial distribution of schools sampled for age or length in the simulated bottom trawl survey and commercial fishery had the greatest influence on effective sample size in the scenarios investigated. Although the aggregation of ages within schools can influence the variability of an estimate compared with simple random sampling (Pennington and Vølstad 1994), rigorous evaluation of effective sample size defined by other methods has not been made previously. Here, effective sample sizes defined by the ratio of empirical to multinomial variance (multinomial approximator, McAllister and Ianelli 1997) or the Dirichlet likelihood (Dirichlet approximator, Williams and Quinn 1998) had a large range of values across age aggregation within schools. In contrast, the sampling approximator (Pennington *et al.* 2002) was similar among the age aggregation scenarios and had great uncertainty in effective sample size. This could be due to large variability in the mean age of the school across replicates and tows, because age–depth preference was implicated in these simulations.

The distribution of schools with respect to depth strata was also influential on the effective sample size of the simulated bottom trawl survey and commercial fishery. For a number of years in both bottom trawl survey and commercial fishery, effective sample size was sometimes larger when schools were distributed randomly across depth strata

than when the mean age of the school increased with depth. The relationship between effective sample size and the spatial distribution of schools was attributable to the underlying age composition of the population, so when fish schools were distributed randomly across depth strata, age classes were sampled more representatively, resulting in an increase in effective sample size.

In terms of accounting for process error, effective sample size was not sensitive to the spatial distribution of schools when depth strata were sampled uniformly. In the bottom trawl survey, shallow and deep depth strata are sampled less frequently than intermediate depths because the survey is designed to obtain abundance estimates for multiple species, not just POP (Hanselman *et al.* 2007). Therefore, effective sample size determined with current methods will be sensitive to the spatial distribution of schools because some depth strata are undersampled, i.e. where schools have a depth preference. In practice, the commercial fishery is not bound to random sampling designs when selecting depth strata in which to operate, so uniform sampling of depths in the commercial fishery is not realistic. Hence, spatial distribution will be an inherent process error in commercial fishery data that should be considered when evaluating effective sample size for stock assessment models.

Because the number of overall tows from which samples were taken for age composition in the bottom trawl survey and the commercial fishery was similar, spatial clustering of schools had a negligible influence on effective sample size. An intuitive interpretation of this result is that spatial clustering ultimately affects the number of tows that catch fish, or the possible number of tows from which samples of fish are obtained

for age or length composition. As the actual sample sizes for age determination in the bottom trawl survey and the commercial fishery are of similar magnitude, the effective sample sizes were similar. There were some notable differences between effective sample sizes of the age composition from the bottom trawl survey and commercial fishery and the length composition of the fishery. These differences can possibly be attributed to the larger number of samples obtained for length composition compared with age composition, as well as the underlying age and length composition of the population. In some years, this difference in actual sample size had an influence on the effective sample size and was related to multimodal age composition in conjunction with a unimodal length composition.

Ageing error is present in all sources of age composition data, and results from this study broaden previous findings that, in some cases, precision in addition to bias is affected when ageing error is present (Coggins and Quinn 1998). When schools consisted of a single age class, the effective sample size increased when ageing error was present compared with when it was not. This is because with ageing error, more age classes were represented in the sampled age composition. For example, if schools are stratified by age, then the age composition from one haul would be different from that of the next haul, but if there is ageing error, then the age composition from the hauls would not look as different. This will create an undetectable bias in the age composition, but the precision will misleadingly look better. In contrast, ageing error decreased the effective sample size in the bottom trawl survey when each school consisted of the true population age composition or when uncertainty can be attributed to observation error only.

In the observation error scenarios, the largest increases in effective sample size were due to increasing the number of tows sampled, rather than increasing the number of fish sampled in each tow. This result supports the results of previous research on effective sample size in which proposals have been made to reduce the haul time of tows in order to increase the number of tows that can be made (Pennington *et al.* 2002, Cerviño and Saborido-Rey 2006). The result also supports the supposition that reducing the number of fish sampled may not have a negative influence on effective sample size (Pennington and Vølstad 1994) and that an increase in the number of tows sampled will improve precision of age or length composition data (Aanes and Pennington 2003). We propose that when possible, sampling designs be developed for age composition sampling that increases the number of tows from which samples are obtained, even if doing so requires reducing the sample size in any particular tow. Also, further studies should be undertaken to understand age aggregation in schools and whether it can be accounted for in stock assessment models.

Subtleties were discovered in this research that have not been identified previously. The aggregation of ages within a school resulted in effective sample sizes that have different relationships with increases in tows or number of fish sampled. The greatest disparity in the relationship between effective sample size and sampling design was the number of ages within the school, either single age or mixed age. For schools consisting of a single age class, there was no clear reason for increases in effective sample size attributable to the number of tows that were sampled or the sample size. Further, the interannual variability for schools conisisting of a single age class was

substantial and difficult to interpret. Reducing the number of tows sampled when schools consisted of a single age decreased the effective sample size, but this was not necessarily the case when schools consisted of fish of mixed age. For example, when schools consisted of fish of mixed age, the general trend was that effective sample size increased when the number of schools sampled for age or length composition also increased. In a bottom trawl survey, when a number of tows that catch fish are not sampled for age or length, a decrease in the number of fish sampled from each tow is proposed, so that the number of tows from which samples are taken can be increased, potentially increasing the precision of age or length data.

In both the multinomial and Dirichlet methods, the true population proportion must be known to approximate effective sample size, which was possible in this study through computer simulation. Unfortunately, however, the true population age or length composition is never known in practice. Approximating effective sample size is possible from raw survey or fishery age/length sample data following the Pennington *et al.* (2002) approach (the sampling approximation method). Further, with the sampling approximation method, the magnitude of effective sample size was comparable with the effective sample size being approximated with the multinomial and Dirichlet methods, although the uncertainty in effective sample size was large with the sampling approximation method compared with the multinomial and Dirichlet methods, and there were no significant differences in effective sample size among the age aggregation scenarios. Therefore, we suggest that further research is warranted to investigate the sensitivity of the Pennington *et al.* (2002) calculations of effective sample size to age

aggregation within schools. Also, further analysis needs to be devoted to evaluating stock assessment model results against values of effective sample size determined from the Pennington *et al.* (2002) calculations.

The Dirichlet likelihood would be a natural likelihood to include in stock assessment models to fit data on biological composition (Schnute and Haigh 2007), and then effective sample size might be estimated as a parameter within the model to provide a better understanding of the actual uncertainty associated with age and length distribution. This approach would be analogous to other approaches in which uncertainty is estimated within stock assessment models (Deriso *et al.* 2007). Results from this study do not provide strict rules on specifying effective sample size in stock assessment models. However, we believe that the results provide guidance on what major factors might influence effective sample size. For example, if the species of interest is well known to school by cohort, this might be a good reason to set effective sample size low for the composition data. Given the nuances of different assessments and surveys, we leave it to the practitioner to apply these results to their specific situation.

1.5 Acknowledgments

This publication is the result of research sponsored by Alaska Sea Grant with funds from the National Oceanic and Atmospheric Administration (NOAA) Office of Sea Grant, Department of Commerce, under grant NA06OAR4170013 (project number R/31–16), and from the University of Alaska with funds appropriated by the state and the Cooperative Institute for Arctic Research/Alaska Fisheries Science Center

(CIFAR/AFSC) fellowship program. We thank Milo Adkison, Brenda Norcross, and Gary Marty for helpful comments and advice, Kalei Shotwell for the bottom trawl survey data and very helpful geographic representations, and Paul Crone and an anonymous referee for valuable comments on an earlier draft of this paper.

1.6 References

Aanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science, 60: 297–303.

Allen, M. J., and Smith, G. B. 1988. Atlas and zoogeography of common fishes in the Bering Sea and northeastern Pacific. US Department of Commerce, NOAA Technical Report, NMFS 66.

Cerviño, S., and Saborido-Rey, F. 2006. Using the bootstrap to investigate the effects of varying tow lengths and catch sampling schemes in fish survey. Fisheries Research, 79: 294–302.

Coggins, L. G., and Quinn, T. J., II, 1998. A simulation study of the effects of aging error and sample size on sustained yield estimates. *In* Fishery Stock Assessment Models, pp. 955–975. Ed. by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, *et al.* Alaska Sea Grant College Program Report AK-SG-98-01, University of Alaska, Fairbanks. 1054 pp.

Crone, P. R., and Sampson, D. B. 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition. *In* Fishery Stock Assessment Models, pp. 355–370. Ed. by F. Funk, T. J. Quinn, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, *et al.* Alaska Sea Grant College Program Report AK-SG-98-01, University of Alaska, Fairbanks. 1054 pp.

Deriso, R. B., Maunder, M. N., and Skalski, J. R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Canadian Journal of Fisheries and Aquatic Sciences, 64: 187–197.

Deriso, R. B., Quinn, T. J., II, and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42: 815–824.

Doubleday, W. G. 1976. A least squares approach to analyzing catch at age data. International Commission for the Northwest Atlantic Fisheries Research Bulletin, 12: 69–81.

Efron, B., and Tibshirani, R. J. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York. 456 pp.

Fournier, D. 1996. An introduction to AD model builder for use in nonlinear modeling and statistics. Otter Research Ltd, Nanaimo, BC, Canada. 194 pp.

Fournier, D., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1195–1207.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunge*. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105–2116.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis, 2nd edn. Chapman and Hall, Boca Raton, FL. 698 pp.

Hanselman, D. H., Heifetz, J., Fujioka, J., Shotwell, S. K., and Ianelli, J. N. 2007. Gulf of Alaska Pacific Ocean perch. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Gulf of Alaska as Projected for 2008, pp. 563–622 . North Pacific Management Council, Anchorage, AK.

Hanselman, D. H., Quinn, T. J., II, Lunsford, C., Heifetz, J., and Clausen, D. M. 2001. Spatial implications of adaptive cluster sampling on Gulf of Alaska rockfish. *In* Proceedings of the 17th Lowell–Wakefield Symposium: Spatial Processes and Management of Marine Populations, pp. 303–325. University of Alaska Sea Grant Program, Fairbanks, AK. 730 pp.

Lo, N. C. H., Jacobsen, L. D., and Squire, J. L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences, 49: 2515–2526.

Love, M. S., Yoklavich, M. M., and Thorsteinson, L. 2002. The Rockfishes of the Northeast Pacific. The University of California Press, Los Angeles. 414 pp.

Manly, B. F. J. 1997. Randomization, Bootstrap, and Monte Carlo Methods in Biology, 3rd edn. Chapman and Hall, London. 488 pp.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54: 284–300.

McGarvey, R., and Pennington, M. 2001. Designing and evaluating length-frequency surveys for trap fisheries with application to the southern rock lobster. Canadian Journal of Fisheries and Aquatic Sciences, 58: 254–261.

Methot, R. D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. American Fisheries Society Symposium, 6: 66–82.

Methot, R. D. 2000. Technical description of the stock synthesis assessment program. NOAA Technical Memorandum, NMFS-NWFSC-43. 56 pp.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Pennington, M., Burmeister, L. M., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin US, 100: 74–80.

Pennington, M., and Vølstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics, 50: 725–732.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York. 542 pp.

R Development Core Team. 2008. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Schnute, J. T., and Haigh, R. 2003. A simulation model for designing groundfish trawl surveys. Canadian Journal of Fisheries and Aquatic Sciences, 60: 640–656.

Schnute, J. T., and Haigh, R. 2007. Compositional analysis of catch curve data, with an application to *Sebastes maliger*. ICES Journal of Marine Science, 64: 218–233.

Schnute, J. T., and Richards, L. J. 1995. The influence of error on population estimates from catch-age models. Canadian Journal of Fisheries and Aquatic Sciences, 52: 2063–2077.

Sigler, M. F., and Fujioka, J. T. 1988. Evaluation of variability in sablefish, *Anoplopoma fimbria*, abundance indices in the Gulf of Alaska using the bootstrap method. Fishery Bulletin US, 86: 445–452.

Williams, E. H., and Quinn, T. J., II, 1998. A parametric bootstrap of catch-age compositions using the Dirichlet distribution. *In* Fishery Stock Assessment Models, pp. 371–384. *Ed.* by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, *et al.* Alaska Sea Grant Program Report AK-SG-98-01. University of Alaska, Fairbanks. 1054 pp.

Table 1.1. Effective sample size simulation cases. Cases considered in the simulation scenarios for observation error and process error.

Observation error scenarios	
Depth cases	
D_0	Resample depths towed in observer data
D_1	Resample depth strata uniformly
Sample size cases	
S_0	Resample number of tows and sample sizes from the fishery or survey data
S_1	Resample half the number of tows with twice the sample size from fishery or survey data
S_2	Resample twice the number of tows with half the sample size from fishery or survey data
S_3	Resample the number of tows with twice the sample size from fishery or survey data
S_4	Resample twice the number of tows with the same sample size from fishery or survey data
Ageing error cases	
E_0	With ageing error
E_1	Without ageing error
Process error scenarios	
Age aggregations and spatial distribution cases	
A_1	Schools consisting of a single age, distributed randomly across depth strata
A_2	Schools consisting of a single age, with school age increasing with depth
A_3	Schools consisting of mixed ages, distributed randomly across depth strata
A_4	Schools consisting of mixed ages, with school age increasing with depth
A_5	Schools consisting of the true population age composition

Table 1.2. Dirichlet method effective sample size. Effective sample size using the Dirichlet method with 95% lower and upper confidence intervals for case S_0.

Year	n_y	A_1	A_2	A_3
Bottom trawl survey (age)				
1984	1 386	7 (4 , 12)	14 (7 , 25)	491 (141 , 1 215)
1987	1 893	7 (4 , 12)	10 (5 , 19)	552 (155 , 1 487)
1990	1 909	7 (4 , 11)	18 (11 , 31)	146 (47 , 517)
1993	1 795	6 (4 , 10)	27 (15 , 47)	136 (44 , 487)
1996	742	5 (2 , 9)	8 (3 , 17)	328 (68 , 847)
1999	974	9 (6 , 16)	9 (5 , 14)	391 (113 , 846)
2003	1 002	8 (5 , 12)	16 (10 , 25)	140 (66 , 344)
2005	1 035	6 (4 , 10)	8 (5 , 12)	124 (55 , 385)
Commercial fishery (age)				
1998	489	6 (4 , 10)	4 (2 , 6)	245 (103 , 492)
1999	397	7 (5 , 13)	9 (7 , 12)	195 (106 , 354)
2000	707	5 (4 , 8)	4 (3 , 6)	165 (105 , 283)
2001	498	5 (4 , 7)	5 (4 , 7)	201 (112 , 376)
2002	388	10 (7 , 14)	12 (9 , 18)	147 (95 , 253)
2004	809	10 (8 , 14)	11 (8 , 14)	109 (71 , 174)
2005	711	7 (6 , 10)	11 (9 , 15)	553 (207 , 1 048)
2006	750	9 (5 , 12)	7 (6 , 10)	561 (272 , 1 052)
Commercial fishery (length)				
1991	2636	56 (32 , 106)	7 (6 , 9)	207 (99 , 443)
1992	2070	49 (26 , 87)	10 (8 , 14)	178 (81 , 393)
1995	990	13 (7 , 27)	2 (2 , 3)	41 (16 , 136)
1996	1947	15 (8 , 24)	15 (11 , 21)	1 017 (341 , 2 270)
1997	3139	19 (10 , 56)	13 (11 , 18)	1 647 (484 , 3 641)

Table 1.2. Continued.

Year	n_y	A_4	A_5
		Bottom trawl survey (age)	
1984	1 386	224 (74 , 607)	1 490 (708 , 2 829)
1987	1 893	131 (67 , 252)	2 092 (1 018 , 4 025)
1990	1 909	312 (147 , 648)	2 065 (1 032 , 3 863)
1993	1 795	604 (223 , 1 476)	1 944 (962 , 3 523)
1996	742	303 (98 , 773)	795 (341 , 1 618)
1999	974	72 (39 , 135)	1 101 (546 , 2 214)
2003	1 002	77 (49 , 121)	1 113 (596 , 2 051)
2005	1 035	56 (36 , 82)	1 118 (616 , 2 187)
		Commercial fishery (age)	
1998	489	59 (38 , 95)	506 (231 , 904)
1999	397	61 (43 , 87)	406 (214 , 773)
2000	707	37 (29 , 47)	761 (416 , 1 349)
2001	498	33 (26 , 42)	513 (272 , 894)
2002	388	43 (33 , 57)	395 (201 , 664)
2004	809	53 (38 , 65)	838 (481 , 1 425)
2005	711	55 (42 , 69)	747 (427 , 1 326)
2006	750	47 (36 , 59)	789 (445 , 1 392)
		Commercial fishery (length)	
1991	2636	86 (69 , 103)	2 459 (1 314 , 4 410)
1992	2070	139 (98 , 196)	1 856 (987 , 3 326)
1995	990	125 (87 , 171)	766 (362 , 1 275)
1996	1947	395 (237 , 658)	1 512 (832 , 2 487)
1997	3139	502 (351 , 710)	2 794 (1 477 , 4 814)

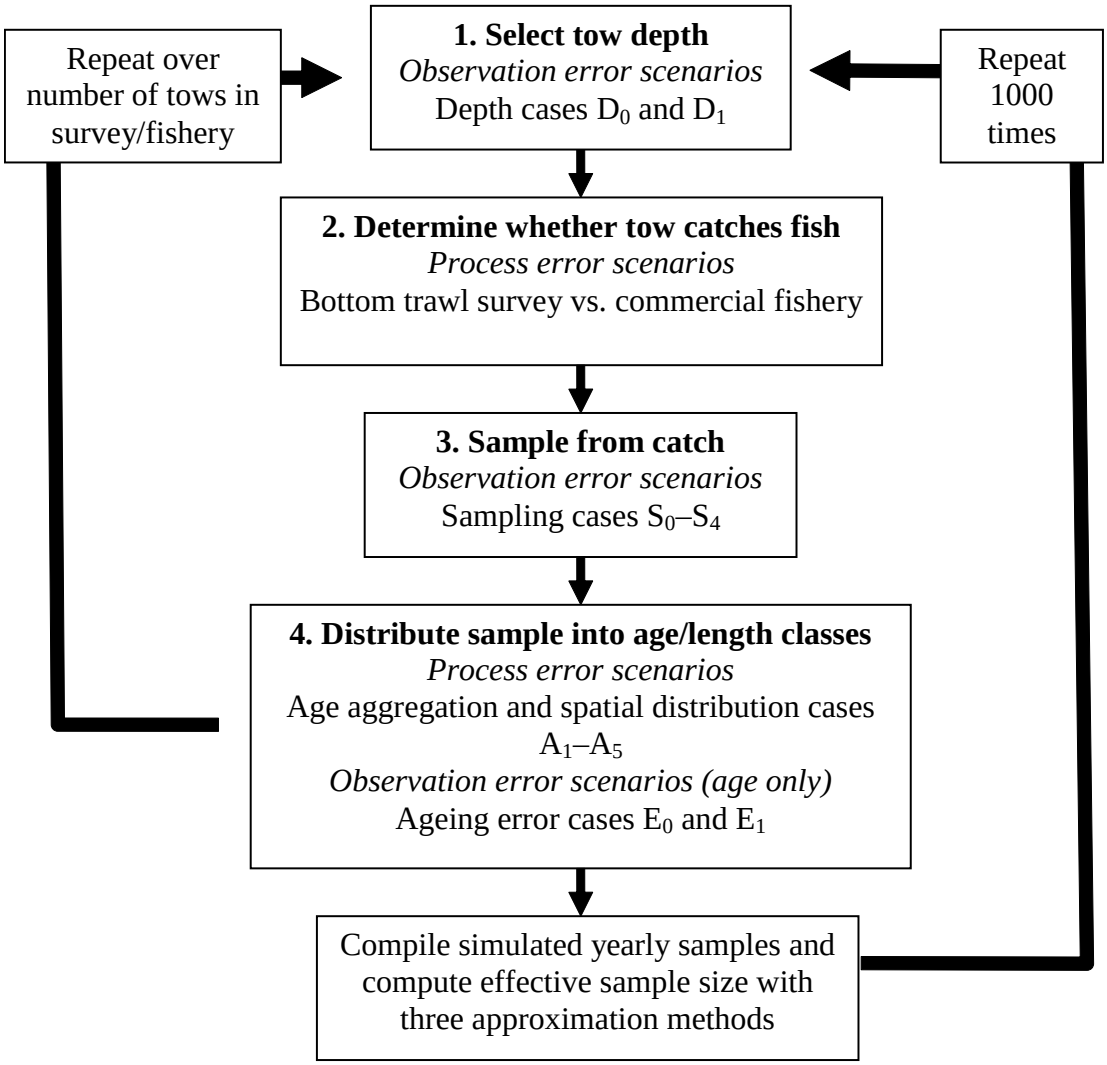

Figure 1.1. Effective sample size scenario cases illustration. Schematic of the methodology and scenario cases to investigate effective sample size associated with age/length composition derived from the Gulf of Alaska Pacific Ocean perch (POP) bottom trawl survey and commercial fishery.

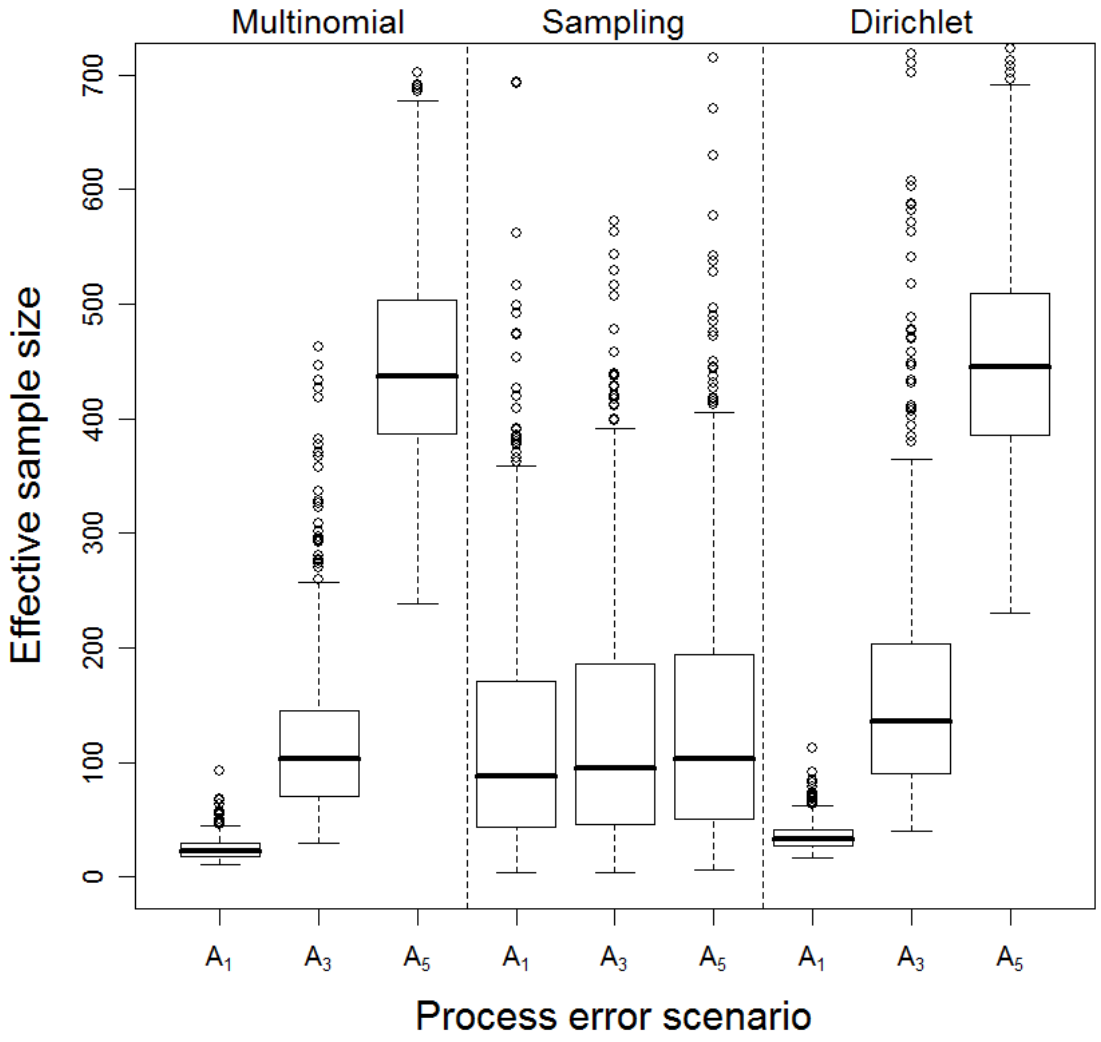

Figure 1.2. Bottom trawl survey effective sample size. Comparison of effective sample size associated with age/length composition based on the multinomial, sampling, and Dirichlet methods for the bottom trawl survey across process error scenarios A_1, A_3, and A_5.

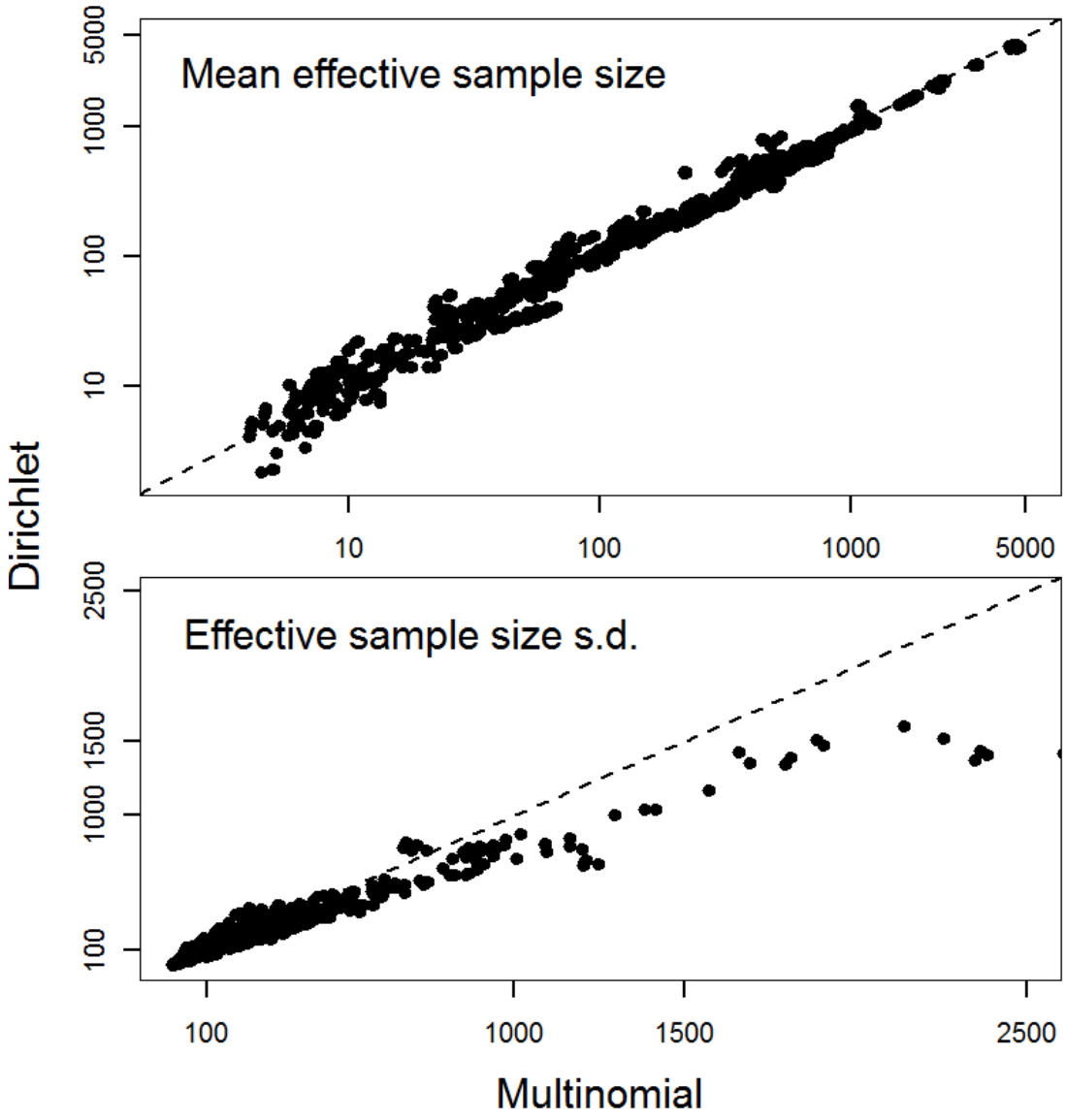

Figure 1.3. Mean and standard deviation of effective sample size. Comparison between the mean (top panel) and standard deviation (s.d., bottom panel) effective sample size estimates from the multinomial and Dirichlet methods for bottom trawl survey and commercial fishery observation error and process error scenarios.

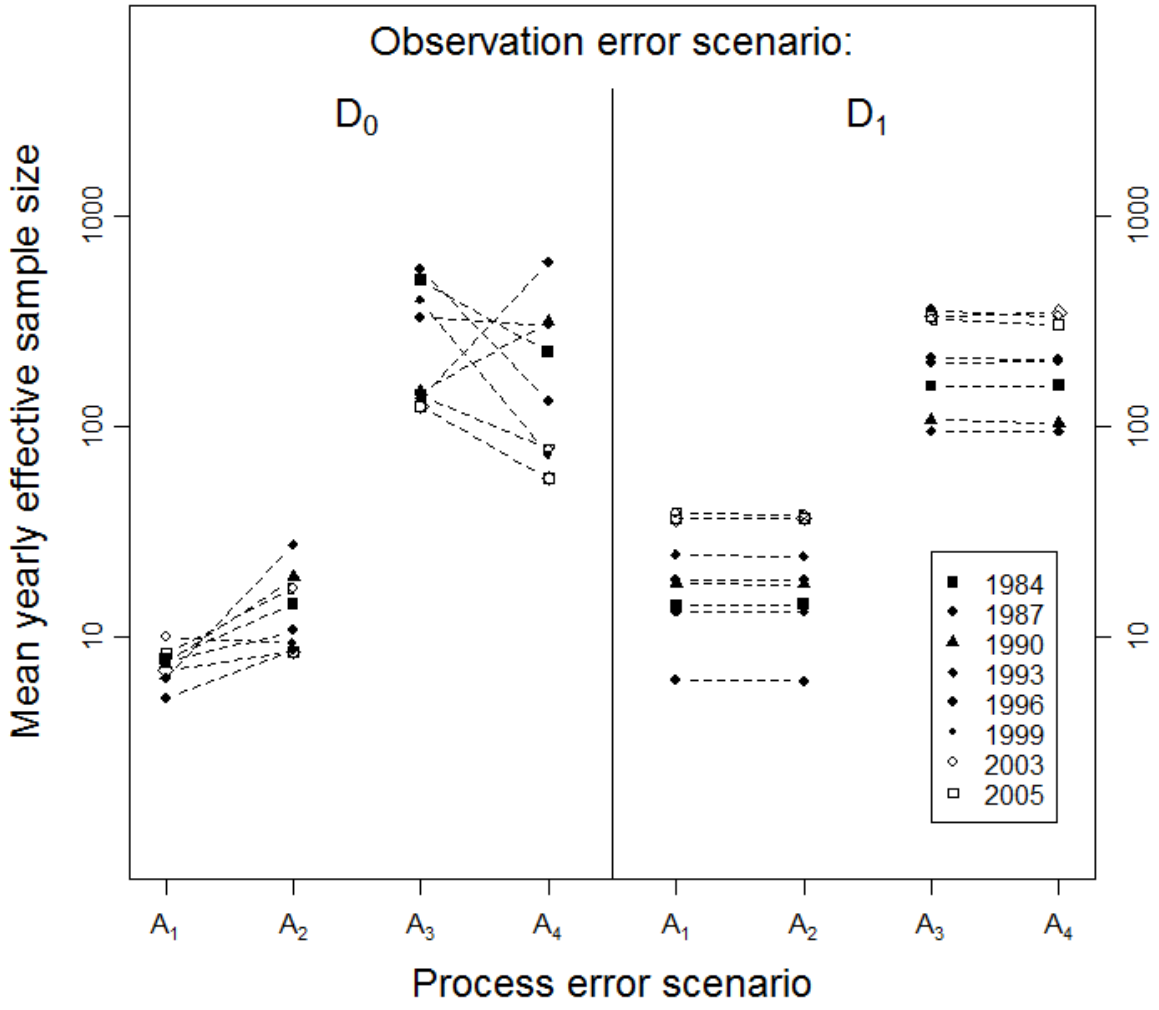

Figure 1.4. Effective sample size by depth case. Bottom trawl survey effective sample size using the Dirichlet method across process error scenarios A_1–A_4 for observation error scenario depth cases D_0 (left panel, depths resampled from observer data) and D_1 (right panel, uniform placement of tows among depth strata). Points connected with dashed lines for illustrative purposes only.

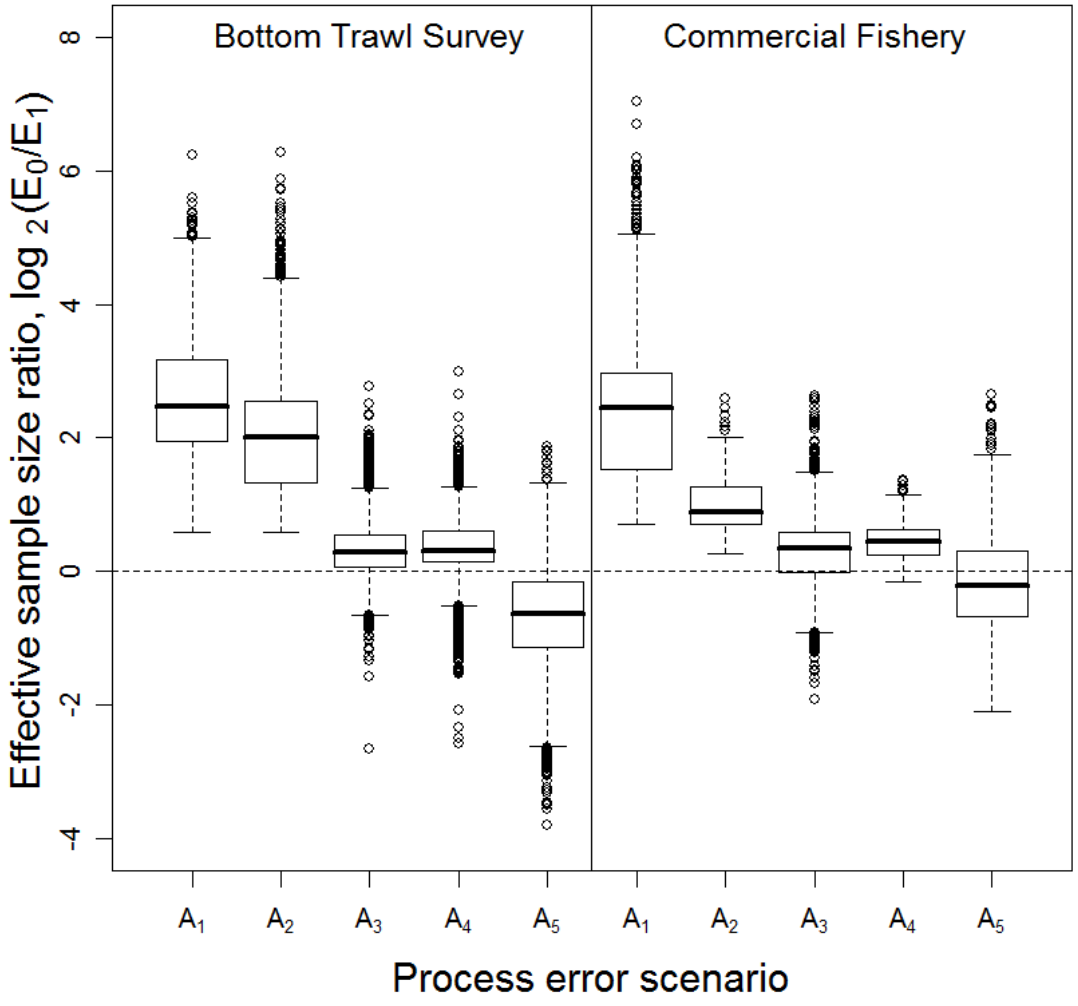

Figure 1.5. Effective sample size ratios by ageing error case. Ratios (log_2) in effective sample size using the Dirichlet method in process error scenarios A_1–A_5 for observation error scenarios with ageing error, case E_0, compared with no ageing error, case E_1, for the bottom trawl survey (left panel) and commercial fishery (right panel).

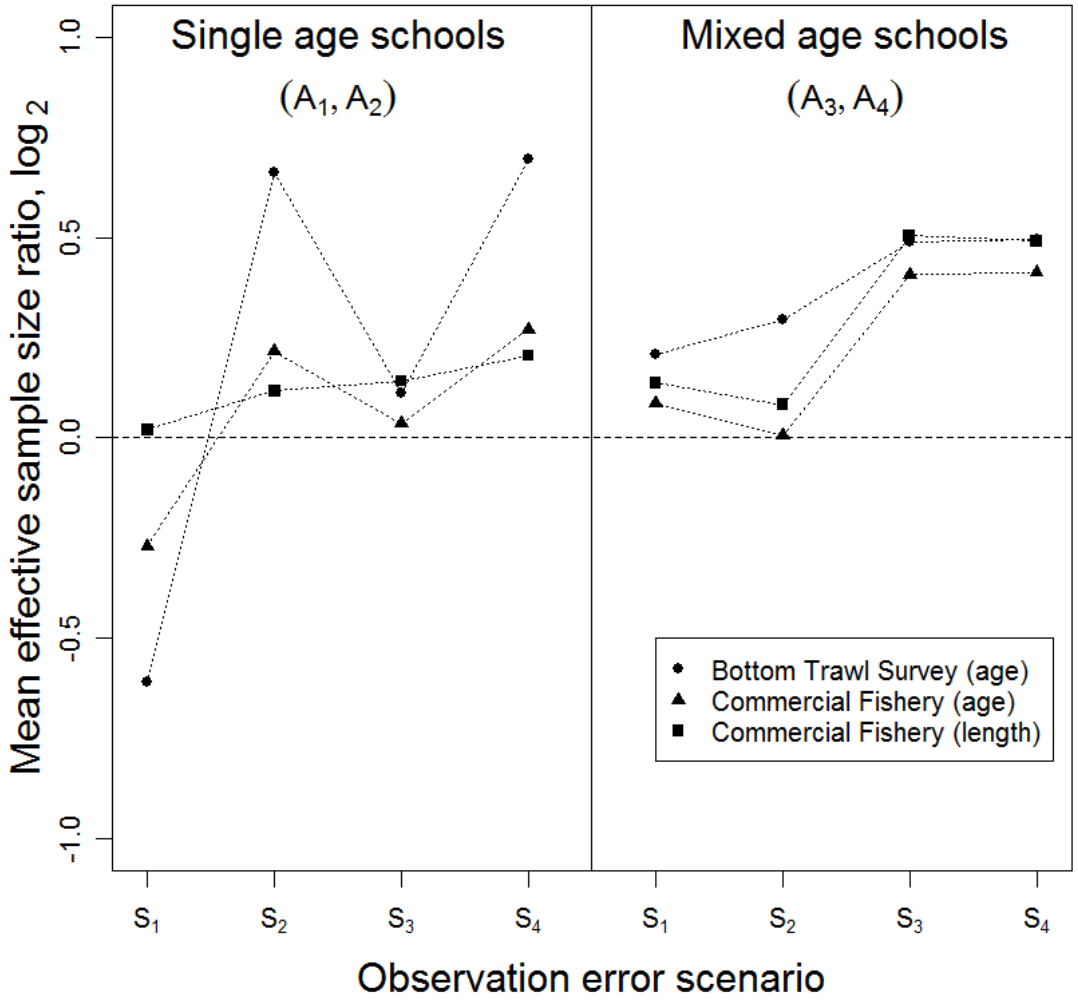

Figure 1.6. Ratios in effective sample size by school age aggregation cases. Effective sample size using the Dirichlet method relative to scenario S_0 from schools of single age (average of process error scenarios A_1 and A_2, left panel) and mixed age (average of process error scenarios A_3 and A_4, right panel) across observation error scenarios for sample size, cases S_1–S_4, for the bottom trawl survey age composition and commercial fishery age and length composition.

Chapter 2: Determining Effective Sample Size in Integrated Age-Structured Assessment Models[1]

Abstract

Age and length composition datasets used in integrated age-structured assessment models commonly exhibit overdispersion. Due to sources of measurement, observation, process, and model specification error, the effective sample size (ESS) is smaller than the actual sample size. In this study we investigated methods to set *a priori* or to estimate ESS when confronted with datasets that include these sources of error, using different statistical distributions and likelihood structures. A number of previously proposed methods to incorporate ESS resulted in accurate estimation of management quantities and parameters when different sources of error were included in the age and length composition data. Three objective methods to incorporate ESS resulted in unbiased management quantities: (1) using sampling theory to derive the ESS from actual age and length composition data, (2) iteratively estimating ESS with the ASA model, and (3) estimating ESS as a parameter with the Dirichlet distribution.

[1] Hulson, P.-J.F., Hanselman, D. H., and Quinn, T. J., II, 2012. Determining effective sample size in integrated age-structured assessment models. ICES Journal of Marine Science 69(2): 281-292.

2.1 Introduction

Developing methods to overcome unknown sources of error has come to the forefront in fishery stock assessment modeling (e.g., Deriso *et al.* 2007), because accuracy and precision of fishery stock assessment model predictions can be extremely sensitive to dataset uncertainty. To quantify sources of uncertainty in fishery stock assessment, most age-structured assessment models (ASA) use the joint negative log-likelihood of the myriad datasets (e.g., Quinn and Deriso 1999). Using the likelihood function requires the specification of dataset uncertainty, a concept often referred to as data weighting (e.g., Deriso *et al.* 1985). The issue of data weighting is not just limited to fishery stock assessment modeling; it applies to any statistical model applied to observed data.

Fundamentally the concept of data weighting is intuitive; the data weight is inversely related to the dataset uncertainty. Under realistic circumstances in data collection for fishery stock assessment, there exist unknown sources of error in datasets that make the issue of data weighting more complicated (e.g., Fournier *et al.* 1990). Fishery stock assessment scientists usually deal with unknown sources of error by inflating the dataset uncertainty through assigning a weighting to a dataset larger than that provided by statistical theory (e.g., Fournier and Archibald 1982, Methot 2000). More recent developments have investigated estimating data weighting (Deriso *et al.* 2007).

Further complication arises with data weighting when there are conflicts between datasets (e.g., NRC 1998, Booth and Quinn 2006). Small changes to the weighting of a dataset can have substantial influence on resulting stock assessment model predictions and inferences. Recent studies deal with data conflicts by examining a range of data

weightings to explicitly identify the influence of dataset conflicts (e.g., Hulson *et al.* 2008).

There are four sources of error inherent to data weighting and subsequent statistical modeling: (1) measurement error, (2) observation error, (3) process error, and (4) model specification error.

We refer to error due to data collection methods as measurement error. Measurement error is often referred to as observation error; herein it is considered a separate category from observation error. Examples of measurement error include ageing error due to reading otoliths or scales, which can be accounted for by applying an ageing error matrix in fishery stock assessment models (e.g., Coggins and Quinn 1998, Hanselman *et al.* 2007), or bias in count data due to diver estimation, which is overcome with diver calibration methods (e.g., Carlile *et al.* 1999).

Observation error's contribution to data weighting is the variance in datasets due to random sampling of a population. Error due to random sampling arises from sub-sampling a population; the variability in the observations can be quantified through the use of sampling theory and would be associated with the sampling design employed (e.g., Thompson 2002).

Error due to unknown changes in population parameters is classified as process error. An example of process error is changes in catchability due to varying spatial distribution of sampling units over time (e.g., Wilberg and Bence 2006).

Model specification error occurs when an incorrect or oversimplified mathematical or statistical model is applied to datasets. An example of this is that in many fish species

schools are composed of cohorts that are close in age, resulting in a sample of age composition that is not representative of the true age composition of the population (Pennington and Vølstad 1994). It is unclear how to model age aggregation within a school that is sampled and in this case not knowing the school structure is treated as model specification error (e.g., McAllister and Ianelli 1997).

Determining dataset weighting in a statistical model becomes more complicated when the dataset's uncertainty contains unknown process and model specification errors. This is because the essential characteristic of both observation and measurement error is that uncertainty can be quantified with statistical theory when fitting a model to data. Thus, sources of process and model specification error cause an underestimation of the total uncertainty in datasets because current statistical theory only considers sources of observation and measurement error to quantify dataset uncertainty (Quinn and Deriso 1999).

ASA models fit two structures of time series for which dataset weighting must be determined: age and length composition datasets, and population index datasets. A number of likelihood functions can be used that are associated with the structure of the datasets (Schnute and Fournier 1980, Williams and Quinn 1998, Schnute and Haigh 2007). The log-normal distribution is often used in ASA models, and a previous study investigated estimating the data weighting in this likelihood function (Deriso *et al.* 2007). Data weighting is further complicated in ASA models because random or independent walks are often employed to estimate annual deviations in parameters (e.g., Wilberg and Bence 2006), a type of process error, and the annual deviations are constrained by using a

normal likelihood in which the deviation uncertainty must be specified. Data weighting in the normal likelihood has received attention in the literature, but estimation of data weighting for non-normal statistical distributions has seldom been evaluated (Schnute and Haigh 2007).

The dataset uncertainty and associated data weighting of age and length composition data is first a function of sample size. To accommodate unknown process or model misspecification error many analysts specify a number for sample size that is smaller than the actual sample size, called the effective sample size (ESS, e.g., McAllister and Ianelli 1997, Hanselman *et al.* 2007). From a sampling theory perspective, ESS is defined as the required number of samples such that the variability from complex sampling designs would be the same as that based on a simple random sample (e.g., Pennington *et al.* 2002). There are a number of proposed methods to fit age or length compositional data while incorporating ESS (e.g., Fournier and Archibald 1982, Methot 2000, also see section 3). However, the effect of process or model specification error on estimates obtained from ASA models with various techniques to specify ESS has rarely been studied. The accuracy of parameter estimation when various ESS techniques are employed has not been investigated. Consequently, stock assessment scientists use differing approaches to set ESS, which has led to inconsistencies among ASA models currently used to manage commercial and recreational fisheries.

In the current study, methods to set *a priori* or to estimate ESS used for weighting age and length composition data within ASA models were investigated with known sources of measurement, observation, process, and model specification error. The data and ASA

model for Gulf of Alaska Pacific Ocean Perch (*Sebastes alutus*, POP) were used because the time series of age and length compositions goes back to the 1960s and the ages fitted in the model include 24 age classes, a complex situation to model (Hanselman *et al.* 2007). The focus of this study was twofold: (1) to evaluate bias in parameter and management quantity estimates obtained from ASA models using various proposed methods to incorporate ESS, and (2) to attempt to identify objective methods to incorporate ESS into ASA models when fitting age and length composition data that contains different sources and magnitudes of error.

2.2 Methods

Estimation methodologies that differed in their treatment of ESS were applied to age and length composition datasets generated by an operating model that included various types of measurement, observation, process, and model specification errors. The operating model and estimation model use the same population dynamics and estimate the same parameters; the primary differences between the operating and estimation model were the statistical distributions used for age and length composition. Previous methods to set ESS *a priori* (e.g., Methot 2000) or to estimate ESS (e.g., Pennington *et al.* 2002) were among those evaluated. Performance and evaluation of the estimation model was based on bias in estimates of parameter values and management quantities compared to 'known' values from the operating model.

2.2.1 Operating model and dataset simulation

In this study, the operating model was based on the ASA model used in the 2007 stock assessment and fishery evaluation (SAFE) for POP. For details the reader is referred to the 2007 SAFE report (Hanselman *et al.* 2007) for equations describing population dynamics and observed datasets. The parameter values obtained and the population dynamics equations used in the 2007 SAFE for POP provided the known population structure in the operating model. Stochastic error was then generated in the population index datasets and age and length composition datasets that were provided from the operating model.

The 2007 ASA model for POP was fit to datasets that included age and length composition and population index data. Age and length compositions were comprised of 24 age classes (ages 2-25) and 23 length bins. As in the 2007 SAFE, commercial fishery age composition was available for 1990, 1998-2002, and 2004-2006; the length composition was available for 1963-1977, and 1991-1997. Also as in the SAFE, the bottom trawl survey (BTS) age composition was collected triennially from 1984-2005. The population index data included time series of observed total fishery yield (1961-2007) and BTS biomass (triennially from 1984-2007).

Population index datasets generated from the operating model included total commercial fishery yield and BTS biomass. For both sets of index data, error was generated in the 'known' values from the SAFE based on standard deviations (SD) used in the data weighting for the 2007 ASA model. For total fishery yield, log-normal error in each year from 1961-2007 was generated with a SD of 0.1. BTS biomass was fitted with

annual estimates of the SD based on the sampling design (Hanselman *et al.* 2007); in this study the SD's were used to generate annual normal error (Table 2.1). Process or model specification error in population index observations was not investigated.

Age and length composition datasets were generated in a stepwise process that was similar to the actual experimental design of the BTS to allow for calculation of ESS though sampling theory (Chapter 1), rather than mimic process or model specification error by generating data with an over-dispersed distribution. The age and length composition data was generated through a four step process. In the first step, the depth stratum of the tow was selected. This was accomplished by resampling the tow depths from observer data of the commercial fishery or BTS. In the second step, the number of fish captured in the tow was produced by generating a binomial random variable based on the observed probability of a tow containing fish; if the tow caught fish then the total catch in the tow was resampled from observer data. In the third step, random tows were examined for age or length measurement using the frequency of sampling by observers, and if a sample was to be taken the sample size was generated by resampling the observer data sample sizes. In the fourth step, the age and length distribution of the sample was generated with a multinomial draw based on the sample size and the underlying age structure of the school.

The data generation scenarios subsequently applied to the estimation model included sources of measurement, observation, process, and model specification error. Given the large number of possible scenarios, we isolated factors that are common to many groundfish assessments to obtain a manageable number. Measurement error was

classified as ageing error when reading otoliths for age composition; ageing error was included in all simulations. Observation error was considered to be the spatial placement of tows, the number of tows from which a sample was obtained, and the sample size of fish taken for age or length composition, which followed from the actual observer data for the commercial fishery and BTS (Chapter 1). Process error was defined as random annual spatial distribution of schools with depth. Model specification error was the age aggregation within the school that was sampled. Three cases for model specification error were considered: schools were composed of completely mixed ages that followed the underlying population age composition (scenario G_1), schools were composed of partially mixed ages (scenario G_2), and schools were composed of only a single age class (scenario G_3). For each data generation scenario, index and age and length composition datasets were replicated 1,000 times (e.g., Efron and Tibshirani 1993). Each estimation model combination was fit to the same generated datasets to allow for direct comparison.

2.2.2 Estimation model

The estimation model was fit to the simulate datasets from the operating model with two classes of weighting methods for the age and length composition datasets: (1) ESS was set *a priori* (ESS_{set}), or, (2) ESS was estimated (ESS_{est}, Figure 2.1). Different statistical distributions and likelihood structures were also investigated for each method of weighting age and length composition. Estimation of dataset weighting for the population index data or parameter deviations was not considered; the SD's used for generating these datasets were treated as known.

Six ESS treatments were considered for the data weighting method that set ESS *a priori* (ESS_{set}, Table 2.2). These treatments were drawn from previous publications that span the last three decades. The second treatment is the most similar to using the actual sample size when fitting age and length composition; ESS was set at the minimum between the actual sample size and 1000 (Fournier *et al.* 1998). The fourth treatment used formulas presented by Pennington *et al.* (2002) to calculate ESS from commercial fishery and BTS age and length samples. The sixth treatment was an ideal case in which we used an approximated value of ESS in which the known underlying age composition and the sampled age composition are used in the Dirichlet distribution that treats ESS as a random variable (Chapter 1). The average actual sample sizes and average ESS from the Dirichlet approximation is shown in Table 2.3.

We also explored six treatments in which ESS was estimated (ESS_{est}) by the ASA model (Table 2.2). The first treatment consisted of an iterative approach that was repeated until the input and output effective sample sizes converged (e.g., McAllister and Ianelli 1997). In the second treatment, ESS was defined by the ratio of theoretical to empirical variance within the ASA model (derived in McAllister and Ianelli 1997). The ASA model estimated ESS as a parameter in treatments three to six. In treatments three and four, mean ESS was estimated across all years for each dataset (e.g., Fournier *et al.* 1990). In treatments five and six, annual parameters for ESS were estimated (e.g., Schnute and Haigh 2007, Williams and Quinn 1998). Treatments four and six incorporated the derivative of the negative log-likelihood with respect to sample size as a constraint function (see derivations in Appendix 2.B, e.g., Deriso *et al.* 2007). For the

normal distribution, the maximum likelihood estimate (MLE) for sample size results in a closed form solution (see derivation in Appendix 2.B). However, a closed form solution for the MLE of ESS is not possible for the Dirichlet, multinomial, and compound multinomial-Dirichlet, so the MLE constraint function was introduced into the ASA model to aid in obtaining convergence.

Under each data weighting method and ESS treatment, four statistical distributions were assumed for age and length composition: the compound multinomial-Dirichlet distribution (C, often referred to as the polya mixture distribution), the Dirichlet distribution (D), the multinomial distribution (M), and the normal distribution (N, Figure 2.1). The compound multinomial-Dirichlet has received little attention in fisheries literature, but has been used in other fields to fit composition data (e.g., Duncan and Wilson 2008). The Dirichlet distribution has received more attention recently as a potential candidate for fitting composition data (e.g., Schnute and Haigh 2007). The multinomial is a commonly used distribution to fit composition data (e.g., Quinn and Deriso 1999). The normal distribution (and related log-normal) is frequently used in ASA models and provides a least-squares method that is often contrasted with the maximum likelihood method (e.g., Fournier and Archibald 1982, Deriso *et al.* 1985).

For computational reasons, three likelihood structures were considered for each statistical distribution: (1) the full likelihood, L_1, (2) the approximate likelihood, L_2, and (3) the normalized likelihood, L_3 (derived in Appendix 2.A, Figure 2.1). The full likelihood did not omit any part of the likelihood. The approximate likelihood omitted constant terms from the likelihood, a common practice in most ASA models (i.e.,

Fournier and Archibald 1982). The normalized likelihood was determined by dividing the full likelihood by the likelihood when the model predicted value equaled the observed value. This is equivalent to subtracting from the negative log-likelihood the value when the predicted equals the observed proportion which is thought to "robustify" the optimization process (Fournier *et al.* 1990). The normal distribution is a special case such that the approximate and normalized likelihood structures are identical (Appendix 2.A). In this study we used the technique presented by Fournier and Archibald (1982) in which the variance in the normal distribution was estimated by using the variance of a proportion, thus, the likelihood accommodates the inclusion of ESS. Theoretically, the different likelihood structures should behave the same (i.e., obtain the same parameter estimates); we considered the different likelihood structures to evaluate what occurs when ESS is treated as a random variable rather than a constant.

2.2.3 Estimation model performance

The metrics used to evaluate model performance encompassed both the capability of the ASA model to fit the generated age and length composition and the resulting estimates of management quantities and parameters. Models for which parameter estimation converged less than 95% of the time were rejected. A subset of parameters and resulting management quantities were selected for investigation and presentation in order to simplify bias results, because over 130 parameters were estimated and the time series of population estimates extends back to 1961. We particularly focused on Acceptable Biological Catch (ABC), the ultimate output in a stock assessment and an indicator of

potential bias in total biomass estimates, especially at the end of the time series. We also emphasized parameters for mean recruitment (*R*), fishing mortality (*F*), and natural mortality (*M*), because these parameters are important for understanding the biological processes and anthropogenic influences affecting total abundance. For ABC, *R*, *F*, and *M* boxplots are shown, where the box represents the inter-quartile range from the estimation models and the whiskers represent the 95% confidence interval. We consider an estimate to be significantly biased if the true value from the operating model is outside the 95% confidence intervals.

In total there were 144 combinations of data weighting methods, ESS treatments, statistical distributions, and likelihood structures applied in the estimation model, hereafter referred to as estimation model combinations (Figure 2.1). Combined with the 3 data generation scenarios, the resulting number of combinations was 432. A stepwise procedure was used to reduce the number of scenarios considered. Initially the estimation model combinations were fitted to data that only contained measurement and observation error (data generation scenario G_1), and estimation model combinations were omitted from further analysis if rejection rates exceeded 5%.

2.3 Results

2.3.1 Data weighting method ESS_{set}

None of the ESS treatments across the statistical distributions and likelihood structures resulted in rejection rates that exceeded 5% when ESS was set *a priori* (data weighting method ESS_{set}) for data generation scenario G_1 (Table 2.4). The range of

rejection rates was between 0% and 0.5%. For each replicated age and length composition dataset the parameter estimates were identical when ESS was input into the ASA model regardless of the likelihood structure employed, whether the full, approximate, or normalized likelihood structure was used.

For data weighting method ESS_{set} under data generation scenario G_1 none of the 24 estimation model combinations resulted in significant bias for ABC; however, there were several estimation model combinations that resulted in significant bias for *R*, *F*, and/or *M* (Figure 2.2). Overall, the only treatment that did not result in significant bias for any of ABC, *R*, *F*, and *M* was treatment IV (Figure 2.2). The same was nearly true for treatment VI, in which the only significant bias resulted in *M* for statistical distribution D. For treatments I, II, and III the only statistical distribution that did not results in significant bias of ABC, *R*, *F*, and *M* was N (Figure 2.2). For each statistical distribution in treatment V significant bias resulted from one or more of estimates of *R*, *F*, and *M*.

Including process and model specification error into the age and length composition under data generation scenario G_2 resulted in only a single estimation model that produced significant bias in ABC, statistical distribution C for treatment V (Figure 2.3). Significant bias in *R* resulted from the 4 estimation model combinations, significant bias in *F* resulted from 13 of the estimation model combinations, and significant bias in *M* resulted for 20 estimation model combinations (Figure 2.3). Overall, there were only 2 estimation model combinations that did not result in significant bias for ABC, *R*, *F*, and *M*, statistical distribution N for treatment II and statistical distribution M for treatment IV.

Increasing the magnitude of process and model specification error in the age and length composition data under data generation scenario G_3 resulted in significant bias in ABC in all estimation model combinations with the exception of 2, statistical distributions D and N for treatment II (Figure 2.4). Significant bias in *R* resulted from 17 estimation model combinations, significant bias in *F* resulted from 15 of the estimation model combinations, and all of the estimation model combinations resulted in significant bias for *M* (Figure 2.4). For this data generation scenario all of the estimation model combinations resulted in significant bias for at least one of ABC, *R*, *F*, and/or *M*.

2.3.2 Data weighting method ESS_{est}

Twenty-seven estimation model combinations out of 72 resulted in rejection rates of 100% and 5 further estimation model combinations had rejection rates that exceeded 5% for data weighting method ESS_{est} under data generation scenario G_1 (Table 2.4). Similar to the results when ESS was set *a priori,* in treatment I all of the likelihood structures resulted in the same parameter estimates. Further results of estimates of ABC, *R*, *F*, and *M* when ESS are shown for treatments, statistical distributions, and likelihood structures that resulted in rejections rates less than 5% in Table 2.4.

When only measurement and observation error only were included in the age and length composition under data generation scenario G_1, significant bias in ABC resulted from 1 out of 11 estimation model combinations for L_1, 7 out of 9 with L_2, and 5 out of 7 with L_3 (top panel, Figure 2.5). Significant bias in ABC only resulted from L_1 for statistical distribution C with treatment II. Bias in ABC resulted from L_2 for statistical

distributions D and N across all ESS treatments (Figure 2.5). For L_3, bias in ABC occurred with statistical distributions C, D, and M for treatment II, and D for treatment IV and VI (Figure 2.5). The same estimation model combinations that resulted in significantly biased estimates of ABC also resulted in significant bias for *R* and *F* (second and third from top panels, Figure 2.5). All of the estimation model combinations resulted in significant bias in *M* with the exceptions of L_1 for statistical distribution D for treatment III and IV and statistical distribution N for treatment IV, and L_3 for statistical distribution D for treatment VI (bottom panel, Figure 2.5). Only 3 out of 27 estimation model combinations resulted in unbiased estimates of all four quantities, ABC, *R*, *F*, and *M*. These estimation model combinations included L_1 for statistical distribution D for treatments III and IV, and L_1 for statistical distribution N for treatment IV (Figure 2.5).

Including process and model specification error into the age and length composition under data generation scenario G_2 resulted in the same estimation model combinations producing bias in ABC (top panel, Figure 2.6) as when only measurement and observation error were included in the age and length composition data. The same estimation model combinations that resulted in significant bias in ABC in this data generation scenario also resulted in significant bias for *R* and *F* (second and third panels from top, Figure 2.6). Additional estimation model combinations that resulted in significant bias in *F* included L_1 for statistical distribution N for treatment IV and statistical distributions D and M for treatment VI, L_2 for statistical distribution M for treatment IV, and L_3 for statistical distribution M for treatments IV and VI (Figure 2.6). Significant bias in *M* for L_1 resulted from statistical distribution C for treatments I and II,

D for treatments III, IV, and V, and M for treatment VI (Figure 2.6). All estimation model combinations resulted in significant bias for *M* with L_2, as did all estimation model combinations for L_3 with the exception of statistical distribution D for treatment IV (Figure 2.6). Only 2 out of 27 estimation model combinations resulted in unbiased estimates of all four quantities ABC, *R*, *F*, and *M* under data generation scenario G_2: L_1 for statistical distributions D and M for treatment I (Figure 2.6).

Significant bias in ABC resulted from all the estimation model combinations under data generation scenario G_3 (Figure 2.7). The same result held for estimates of *R* (second from top panel, Figure 2.7). The only estimation model combination that did not result in significant bias for *F* was L_3 for statistical distribution M for treatment II, although the variability in *R* was large (third from top panel, Figure 2.7). The only estimation model combinations that did not result in significant bias in *M* was L_2 and L_3 for statistical distribution D for treatment II, although the variance in *M* was large (bottom panel, Figure 2.7).

2.4 Discussion

In this study we have shown that there exist several ways to statistically model age and length composition and to obtain accurate estimates of management quantities in ASA models. However, there was not a single data weighting method, statistical distribution, or ESS treatment that resulted in unbiased estimates of parameters for all the data generation scenarios. There were a number of estimation model combinations that performed well for all statistical distributions, but the normal distribution was the most

robust when setting ESS *a priori* and the Dirichlet distribution when ESS was estimated. The multinomial distribution also performed well for a number of ESS treatments, but it was very difficult to estimate ESS with the multinomial distribution. The compound multinomial-Dirichlet distribution was a promising distribution to use for age and length composition data; however, implementing this distribution turned out to be very computationally intensive and it didn't perform better than the other distributions. The analysis provided in this study provides guidance to stock assessment scientists when faced with modeling age and length composition data. The estimation model combinations that resulted in unbiased estimates in this study can be incorporated into current ASA models and evaluated based on the characteristics (i.e., age aggregation, etc.) of the particular species modeled.

Biased estimates of management quantities and parameters were rare when only measurement and observation errors were included in the generated age and length composition data. Whereas, when process and model specification errors were also included, significant bias occurred more often. There were several ESS treatments and statistical distributions that resulted in accurate estimates of ABC when process and model specification error were included, although biased parameter estimates for *R*, *F*, and *M* could still result. In the extreme case of model specification error, when schools were composed of only a single age class, accurate estimates of management quantities and parameters were not obtained, whether setting ESS *a priori* or estimating ESS, exposing a potential limitation of ASA models. When schools were composed of a single age class the general pattern across estimation model combinations was that *R* was over-

estimated and *M* was under-estimated, resulting in an ABC that was larger than the true value. There appears to be a very limited amount of information in the age and length composition obtained from sampling age aggregated fish schools to provide accurate estimation *R* and *M*. Under these circumstances the stock assessment scientist should be very cautious when applying an ASA model to estimate the population dynamics. Further research is warranted to find if methods can be developed that overcome extreme age aggregation including studies of school structure and subsequent implementation into ASA models.

Three treatments that objectively implemented ESS into the ASA model resulted in accurate estimates of ABC, although in some instances bias in parameters could result. The first method was to calculate ESS *a priori* from the actual observer data after using sampling theory to provide an estimate of ESS (e.g., Pennington *et al.* 2002). Previous research on using sampling theory to calculate ESS has focused mostly on improving sampling design (e.g., Cerviño and Saborido-Rey 2006), although this method could easily be extended and implemented into ASA models. The second method to objectively implement ESS was to iteratively estimate ESS with the ASA model (e.g., McAllister and Ianelli 1997). Iterative estimation was the only treatment to result in unbiased estimates of ABC, *R*, *F*, and *M* when process and model specification errors were included in the age and length composition and was seen with both the Dirichlet and multinomial distributions. A drawback of iterative techniques is the computational time needed to estimate ESS. The third method to objectively implement ESS into ASA models was to estimate ESS as a parameter. The only distribution that was able to estimate ESS as a

parameter without the MLE constraint function was the Dirichlet distribution. Using this method, fewer biased estimates resulted when ESS was estimated as a mean parameter across years rather than annually, possibly signaling a problem of over-parameterization.

Estimating ESS, whether as a sampling statistic, iterative estimate, or parameter, has rarely been applied in ASA modeling and presents some interesting implications. The main assumption of the sampling statistic approach is that ESS can be inferred directly from the experimental design. An attribute of the iterative estimate of ESS is that the dataset uncertainty defined *a priori* is the same dataset uncertainty that subsequently results from the ASA model. While both of these methods essentially consider ESS to be a random variable, neither provides quantification of the uncertainty in the ESS value used in the ASA model. Estimating ESS as a parameter would aid in the evaluation of the uncertainty in model predictions, as there is an uncertainty term connected with the ESS parameter. This could possibly provide a more complete posterior distribution of the ASA model estimates that takes the underlying uncertainty in the age and length composition data into account.

As most previous studies have shown (e.g., Fournier and Archibald 1982, Methot 2000, Hanselman *et al.* 2007), parameter estimates were not sensitive to the likelihood structure when setting ESS *a priori*, but the likelihood structure did have an effect on the accuracy of management quantities and parameter estimates when estimating ESS. This result is not the same when ESS is estimated as a parameter; the likelihood structure did have an effect on the parameter estimates in the ASA model. The full likelihood structure resulted in fewer biased estimates than the approximate or normalized likelihood

structures. The differences between the likelihood structures when estimating ESS is a computational problem, in theory, the likelihood structure should not matter. Upon parameterizing ESS to fit age and length composition data, an analyst should use the full likelihood structure and not omit any part of the probability density function, which essentially eliminates *ad hoc* omission of certain components of the likelihood function.

The method used here to generate error in age and length composition isolated sources of observation, measurement, process, and model specification error, but did not exhaustively include all sources of error that may be present in actual data collection for age and length composition. We did attempt to include error that most adequately represented what could occur in real world circumstances; however, giving specific recommendations for how to implement ESS into ASA models is difficult due to our simplified representation of the data generation. However, we have been able to identify ESS methods and treatments that do not perform well, even under these simplified cases. The most disconcerting was the performance of defining ESS as the square root of the actual sample size, which is the current method employed to fit POP age and length composition data (Hanselman *et al.* 2007). Even with only measurement and observation error in the age and length composition data, biased estimates of ABC could be obtained. Further, for each statistical distribution tested, parameter estimates for *R*, *F*, and *M* were frequently biased. Bias in ABC, *R*, *F*, and *M* became more prevalent as process and model specification error were also included.

Another simplification that was made in the data generation process was to not include process and model specification error in the population index data (commercial

fishery yield and BTS biomass) and to treat the uncertainty in index data as known. Further research should be conducted to evaluate the effects of process and model specification error on index data while estimating dataset weighting within ASA models and how this relates to the various ESS treatments.

Recent developments in ASA modeling have focused on constructing more complicated types of ASA models, including spatially-explicit models (e.g., Maunder 2001) and ecosystem level models (e.g., Quinn and Collie 2005). In the case of spatially-explicit models, for example, the dividing of age and length composition into spatially-explicit regions creates complications. Following the variance of a proportion, the variance in spatially-explicit age and length composition data would be larger than spatially-integrated age and length composition data. The effect on ESS, however, is unclear. The objective methods for evaluating ESS we have presented are still applicable to these types of situations in ASA modeling, regardless of the spatial and temporal structure of the ASA model. We have shown that there could potentially be a number of methods to implement ESS into statistical modeling of age and length composition that can provide accurate estimates of management quantities and parameters. These methods could be extended to accommodate a large number of other ASA model structures.

2.5 Acknowledgments

This publication is the result of research sponsored by Alaska Sea Grant with funds from the National Oceanic and Atmospheric Administration Office of Sea Grant, Department of Commerce, under grant no. NA06OAR4170013 (project number R/31-

16), and from the University of Alaska with funds appropriated by the state and the Cooperative Institute for Arctic Research/Alaska Fisheries Science Center (CIFAR/AFSC) Fellowship program. We would like to thank Drs. Milo Adkison, Brenda Norcross, and Gary Marty for helpful comments and advice.

2.6 References

Bernardo, J. M. 1976. Algorithm AS 103: Psi (Digamma) Function. Journal of the Royal Statistical Society, Series C, 25(3): 315–317.

Booth, A. J., and Quinn, T. J., II, 2006. Maximum likelihood and Bayesian approaches to stock assessment when data are questionable. Fisheries Research, 80: 169–181.

Carlile, D. W., Larson, R. C., and Hebert, K. P. 1999. Stock assessments of Southeast Alaska herring in 1997 and forecasts for 1998 abundance. Alaska Department of Fish and game, Division of Commercial Fisheries, Regional Information Report 1J99-10, Juneau, Alaska.

Cerviño, S. and Saborido-Rey, F. 2006. Using the bootstrap to investigate the effects of varying tow lengths and catch sampling scemes in fish survey. Fisheries Research, 79: 294–302.

Coggins, L. G., and Quinn, T. J., II, 1998. A simulation study of the effects of aging error and sample size on sustained yield estimates. *In* Fishery Stock Assessment Models, pp. 955–975. Ed. by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, *et al.* Alaska Sea Grant College Program Report AK-SG-98-01, University of Alaska, Fairbanks. 1054 pp.

Deriso, R. B., Quinn, T. J., II, and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42(4): 815–824.

Deriso, R. B., Maunder, M. N., and Skalski, J. R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Canadian Journal of Fisheries and Aquatic Sciences, 64(2): 187–197.

Duncan, K. A., and Wilson, J. L. 2008. A Multinomial-Dirichlet Model for Analysis of Competing Hypotheses. Risk Analysis, 28(6): 1699–1709.

Efron, B., and Tibshirani, R. J. 1993. An introduction to the Bootstrap. Chapman and Hall, New York. 456 pp.

Fournier, D., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39(8): 1195–1207.

Fournier, D., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). Canadian Journal of Fisheries and Aquatic Sciences, 47: 301–317.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South

Pacific albacore, *Thunnus alalunge*. Canadian Journal of Fisheries and Aquatic Sciences, 55(9): 2105–2116.

Hanselman, D. H., Heifetz, J., Fujioka, J., Shotwell, S. K., and Ianelli, J. N. 2007. Gulf of Alaska Pacific ocean perch. *In* Stock Assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2008, pp.563–622. North Pacific Management Council, Anchorage, AK.

Hulson, P.-J. F., Miller, S. E., Quinn, T. J., II, Marty, G. D., Moffit, S. D., and Funk, F. 2008. Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. ICES Journal of Marine Science, 65: 25–43.

Maunder, M. 2001. Integrated Tagging and Catch-at-Age Analysis (ITCAAN): Model development and simulation testing. *In* Spatial Processes and Management of Fish Populations, pp. 123–146. Ed. by G. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and D. Witherell. Alaska Sea Grant College Program No. AK-SG-01-02, Univ. of Alaska, Fairbanks, AK.

Methot, R. D. 2000. Technical description of the stock synthesis assessment program. NOAA Technical Memorandum NMFS-NWFSC-43. 56 pp.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54(2): 284–300.

National Research Council (NRC). 1998. Improving Fish Stock Assessments. National Academy Press, Washington, DC.

Pennington, M., and Vølstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics, 50(3): 725–732.

Pennington, M., Burmeister, L. M., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fisheries Bulletin, 100: 74–80.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative fish dynamics. Oxford University Press, New York. 542 pp.

Quinn, T. J., II, and Collie, J. S. 2005. Sustainability in single-species population models. Philosophical Transactions of the Royal Society B, 360: 147–162.

Schnute, J. T., and Haigh, R. 2007. Compositional analysis of catch curve data, with and application to *Sebastes maliger*. ICES Journal of Marine Science, 64(2): 218–233.

Schnute, J. T., and Fournier, D. 1980. A new approach to length-frequency analysis: growth structure. Canadian Journal of Fisheries and Aquatic Sciences, 37(9): 1337–1351.

Thompson, G. G. 1995. Pacific cod. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions as projected for 1996, pp. 2-1 – 2-28. North Pacific Fishery Management Council, Anchorage, AK.

Thompson, S. K. 2002. Sampling. John Wiley and Sons, Inc., New York. 400 pp.

Wilberg, M. J., and Bence, J. R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences, 63: 2275–2285.

Williams, E. H., and Quinn, T. J., II, 1998. A parametric bootstrap of catch-age compositions using the Dirichlet distribution. *In* Fishery stock assessment models, pp. 371-384. Ed. by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, and C. I. Zhang. Alaska Sea Grant Program Report No. AK-SG-98-01, University of Alaska Fairbanks, AK. 1054 pp.

Table 2.1. Standard Errors (SE) used in data generation of Bottom Trawl Survey (BTS) biomass.

Year	Standard Error
1984	66274
1987	45575
1990	33926
1993	102748
1996	233580
1999	388587
2001	249104
2003	72016
2005	146602
2007	114172

Table 2.2. Description of effective sample size (ESS) treatments evaluated.

Data Weighting Method	ESS treatment	Description	Reference
ESS_{set}: Set *a priori*	I	400 for all years of the commercial fishery and BTS age and length composition	Fournier and Archibald 1982
	II	The minimum between the actual sample size or 1000	Fournier *et al.* 1998
	III	200 for all years of the commercial fishery and BTS age and length composition	Methot 2000
	IV	Sampling theory approximation method when generating age and length composition	Pennington and Vølstad 1994, Pennington *et al.* 2002, Hulson *et al.* 2011
	V	The square root of the actual sample size	Hanselman *et al.* 2007, Thompson 1995
	VI	Dirichlet approximation method when generating age and length composition	Hulson *et al.* 2011
ESS_{est}: Estimate	I	Iterative estimation	McAllister and Ianelli 1997
	II	Annual ESS defined as the ratio of empirical to theoretical variance	McAllister and Ianelli 1997
	III	Mean ESS across years estimated as a parameter within the ASA model	Fournier *et al.* 1990
	IV	Mean ESS across years estimated as a parameter within the ASA model that includes the MLE constraint function	Deriso *et al.* 2007
	V	Annual ESS estimated as a parameter within the ASA model	Schnute and Haigh 2007, Williams and Quinn 1998
	VI	Annual ESS estimated as a parameter within the ASA model that includes the MLE constraint function	Deriso *et al.* 2007

Table 2.3. Mean effective sample size with 95% lower and upper confidence intervals from the Dirichlet approximator.

Year	n_y	G_1	G_2	G_3
		Bottom trawl Survey (age)		
1984	1386	1490 (708,2829)	491 (141,1215)	7 (4,12)
1987	1893	2092 (1018,4025)	552 (155,1487)	7 (4,12)
1990	1909	2065 (1032,3863)	146 (47,517)	7 (4,11)
1993	1795	1944 (962,3523)	136 (44,487)	6 (4,10)
1996	742	795 (341,1618)	328 (68,847)	5 (2,9)
1999	974	1101 (546,2214)	391 (113,846)	9 (6,16)
2003	1002	1113 (596,2051)	140 (66,344)	8 (5,12)
2005	1035	1118 (616,2187)	124 (55,385)	6 (4,10)
		Commercial Fishery (age)		
1998	489	506 (231,904)	245 (103,492)	6 (4,10)
1999	397	406 (214,773)	195 (106,354)	7 (5,13)
2000	707	761 (416,1349)	165 (105,283)	5 (4,8)
2001	498	513 (272,894)	201 (112,376)	5 (4,7)
2002	388	395 (201,664)	147 (95,253)	10 (7,14)
2004	809	838 (481,1425)	109 (71,174)	10 (8,14)
2005	711	747 (427,1326)	553 (207,1048)	7 (6,10)
2006	750	789 (445,1392)	561 (272,1052)	9 (5,12)
		Commercial Fishery (length)		
1991	2636	2459 (1314,4410)	207 (99,443)	56 (32,106)
1992	2070	1856 (987,3326)	178 (81,393)	49 (26,87)
1995	990	766 (362,1275)	41 (16,136)	13 (7,27)
1996	1947	1512 (832,2487)	1017 (341,2270)	15 (8,24)
1997	3139	2794 (1477,4814)	1647 (484,3641)	19 (10,56)

Table 2.4. Rejection rates for data generation scenario G_1. Shown across effective sample size (ESS) treatments for each statistical distribution (C: Dirichlet compound multinomial, D: Dirichlet, M: multinomial, N: normal) and likelihood structure (L_1: Full likelihood, L_2: Approximate likelihood, L_3: Normalized likelihood).

ESS treatment	Statistical Distribution	ESS_{set}	ESS_{est}		
			L_1	L_2	L_3
I	C	0.1%	0.0%		
	D	0.0%	0.1%		
	M	0.1%	0.7%		
	N	0.0%	**24.8%**		
II	C	0.0%	0.3%	**100.0%**	0.1%
	D	0.1%	0.2%	0.1%	0.1%
	M	0.0%	**7.1%**	**6.3%**	0.6%
	N	0.1%	**100.0%**	**11.0%**	
III	C	0.1%	**100.0%**	**100.0%**	**100.0%**
	D	0.3%	0.1%	3.1%	**100.0%**
	M	0.3%	**100.0%**	**100.0%**	**100.0%**
	N	0.1%	**100.0%**	**100.0%**	
IV	C	0.1%	**100.0%**	**100.0%**	**100.0%**
	D	0.2%	0.0%	0.1%	0.0%
	M	0.2%	**37.2%**	0.2%	0.7%
	N	0.5%	0.0%	0.5%	
V	C	0.0%	**100.0%**	**100.0%**	**100.0%**
	D	0.0%	0.2%	0.3%	**100.0%**
	M	0.1%	**100.0%**	**100.0%**	**100.0%**
	N	0.1%	**100.0%**	**100.0%**	
VI	C	0.0%	**100.0%**	**100.0%**	**100.0%**
	D	0.0%	1.1%	0.2%	0.5%
	M	0.1%	0.1%	0.3%	0.0%
	N	0.5%	**100.0%**	1.1%	

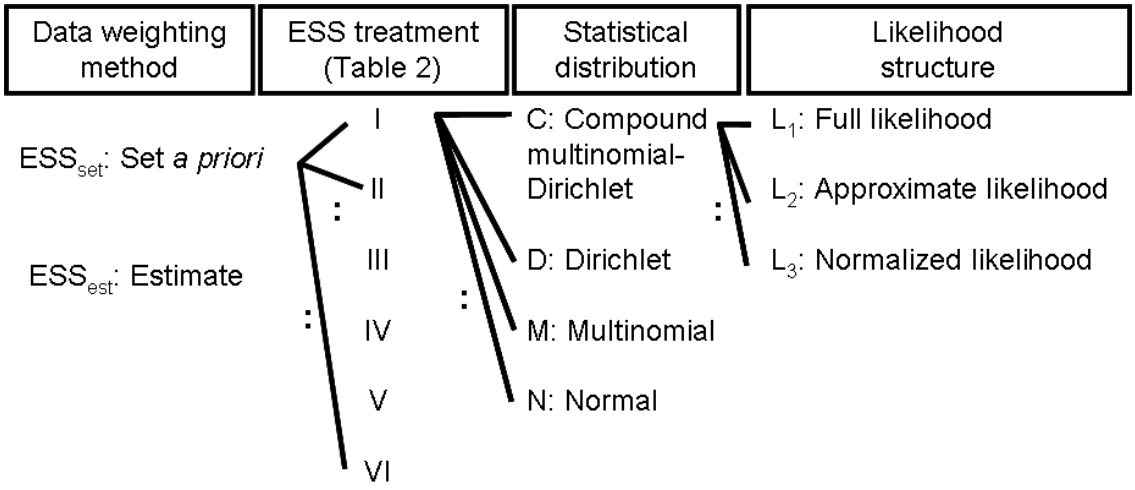

Figure 2.1. Model scenarios considered for estimation of effective sample size. Schematic of the data weighting methods, effective sample size (ESS) treatments, statistical distributions, and likelihood structures considered in the estimation model resulting in 144 combinations applied to each of the three data generation scenarios.

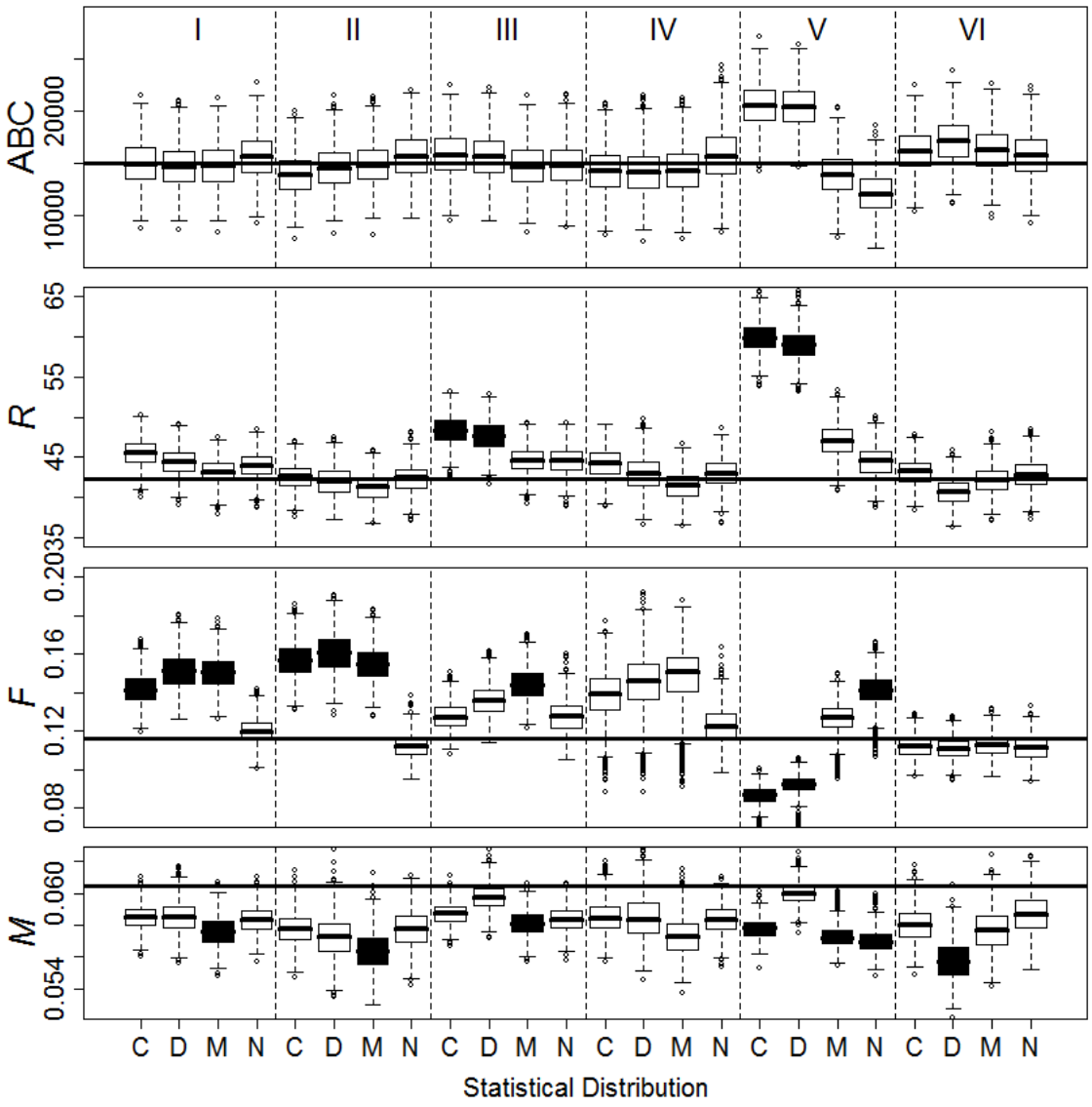

Figure 2.2. ESS_{set} boxplots for data generation scenario G_1. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) across estimation model combinations (C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). Boxes with a black background represent estimates resulting in significant bias for which the 95% confidence intervals do not contain the true value (solid horizontal line).

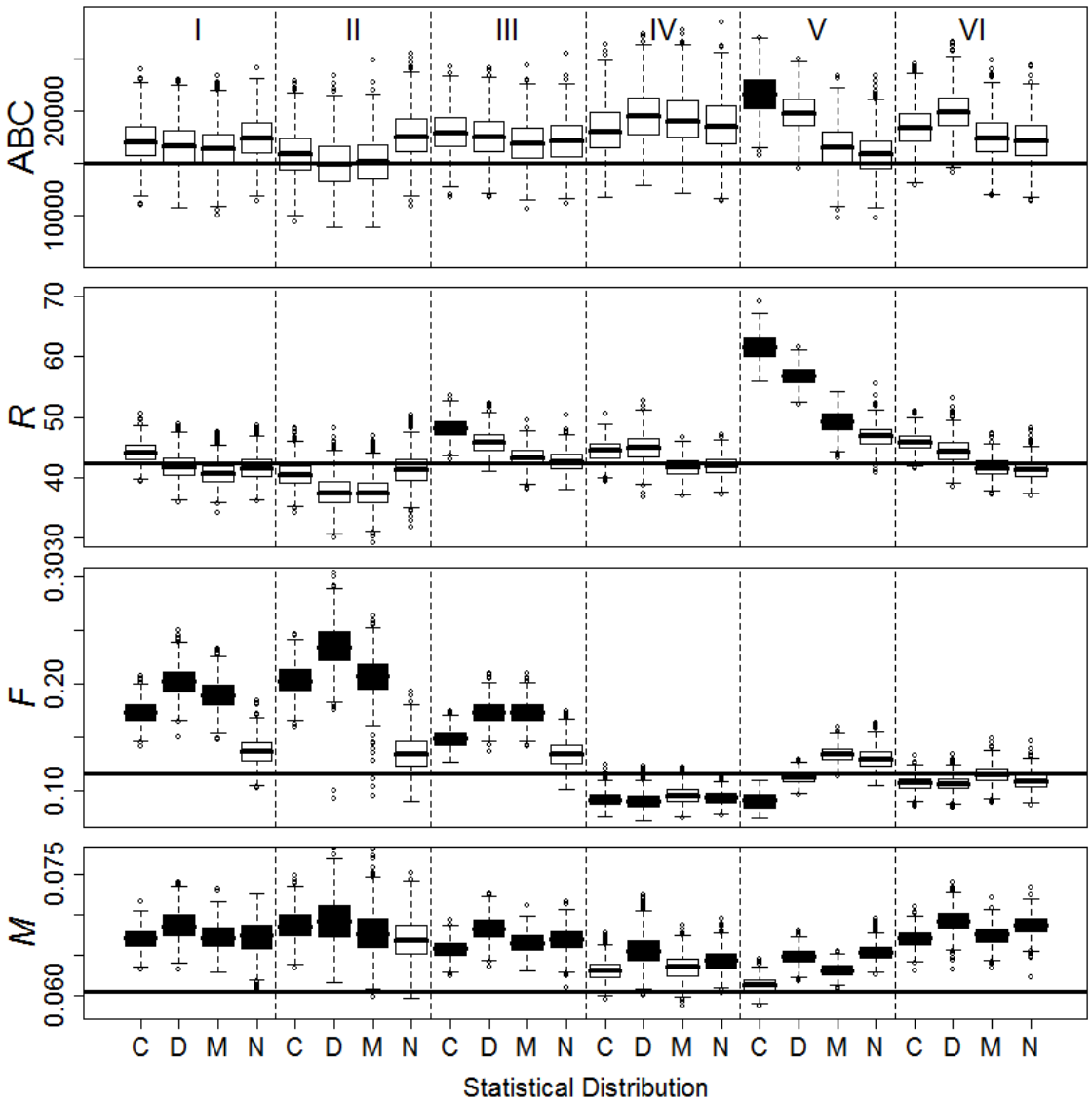

Figure 2.3. ESS_{set} boxplots for data generation scenario G_2. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) across estimation model combinations (C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). Boxes with a black background represent estimates resulting in significant bias for which the 95% confidence intervals do not contain the true value (solid horizontal line).

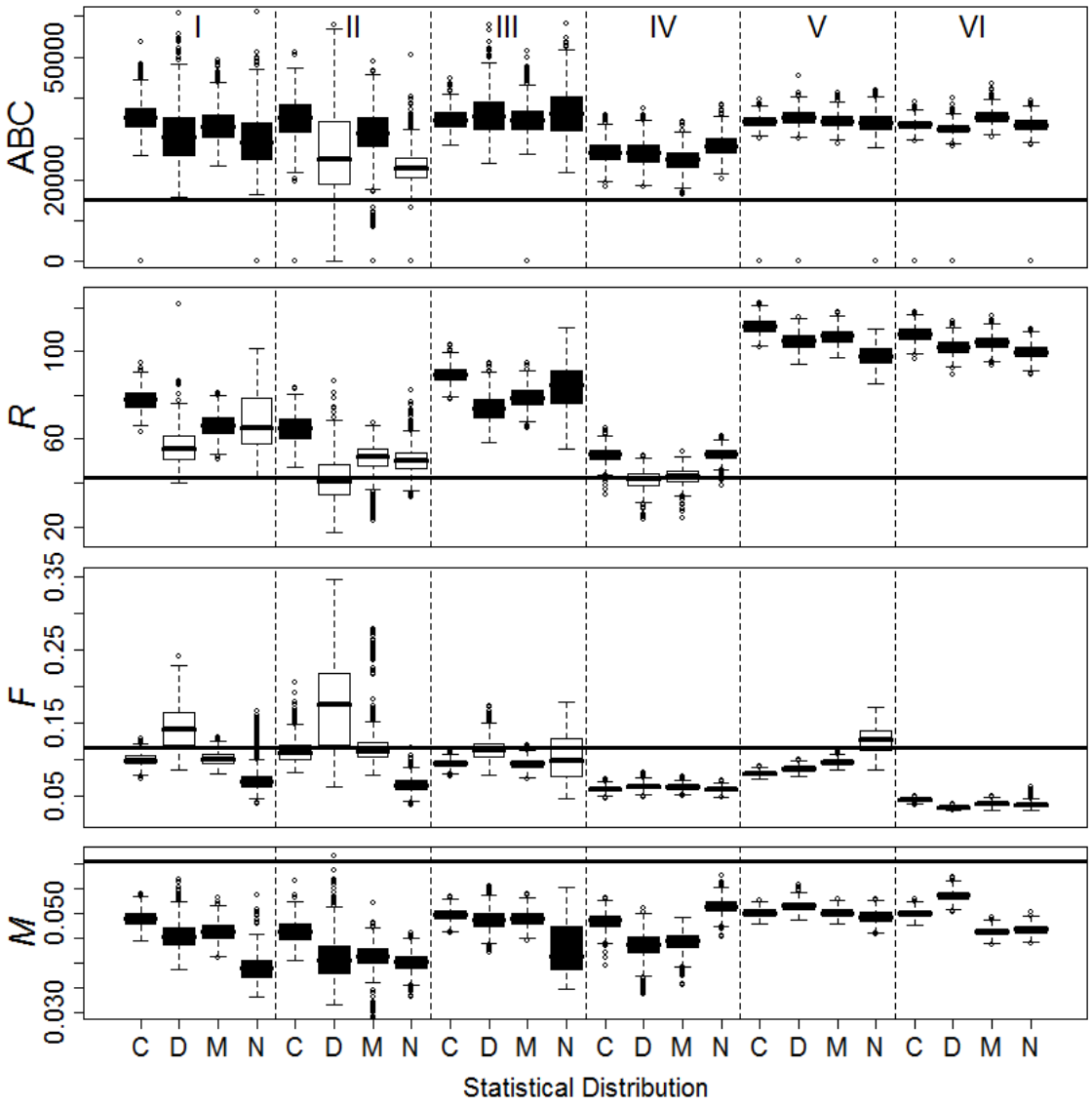

Figure 2.4. ESS_{set} boxplots for data generation scenario G_3. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) for across estimation model combinations (C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). Boxes with a black background represent estimates resulting in significant bias for which the 95% confidence intervals do not contain the true value (solid horizontal line).

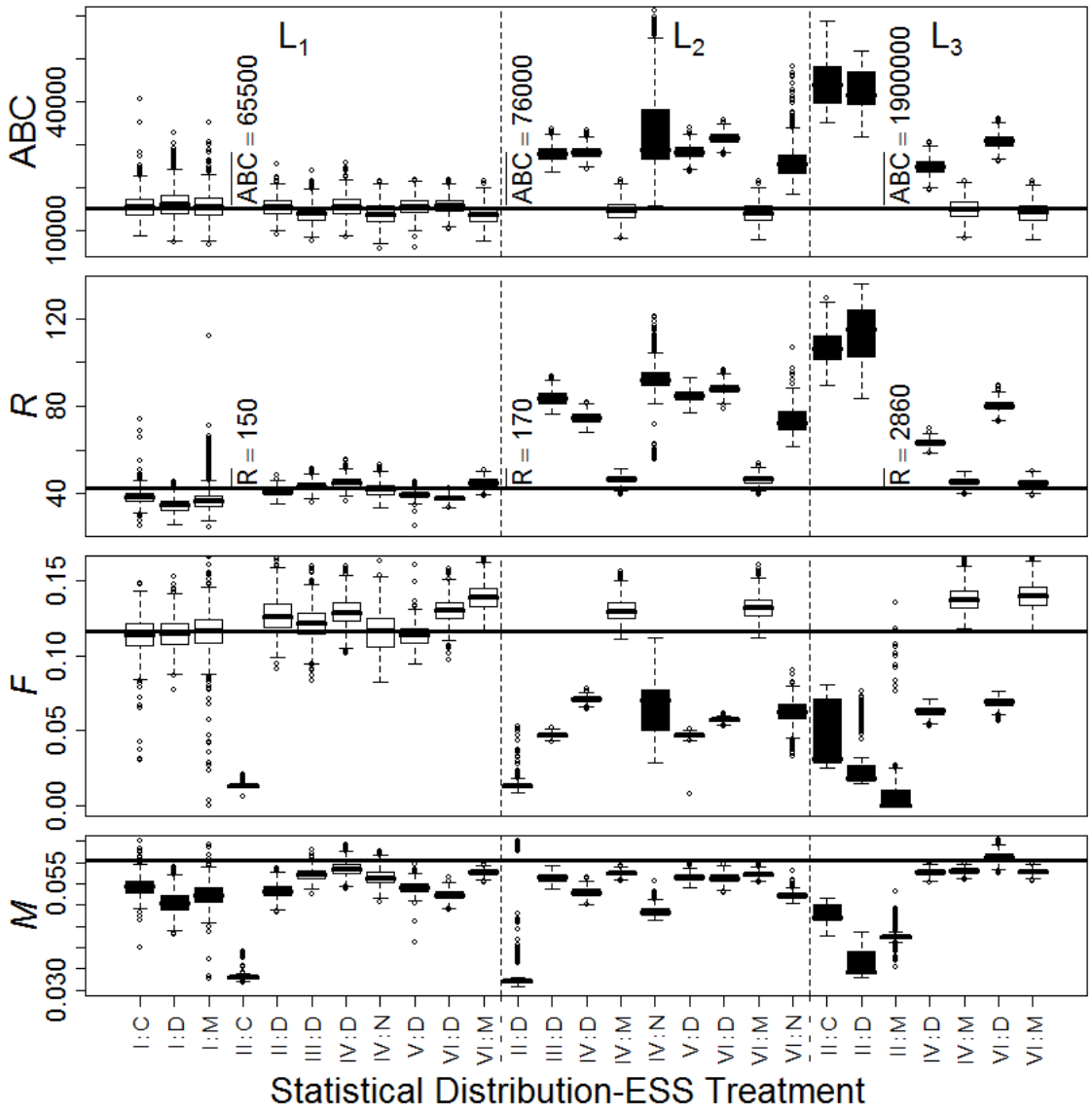

Figure 2.5. ESS_{est} boxplots for data generation scenario G_1. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) across estimation model combinations (L_1: Full likelihood structure, L_2: Approximate likelihood structure, L_3: Normalized likelihood structure, C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). For results beyond the range of the *y*-axis the mean is shown in text.

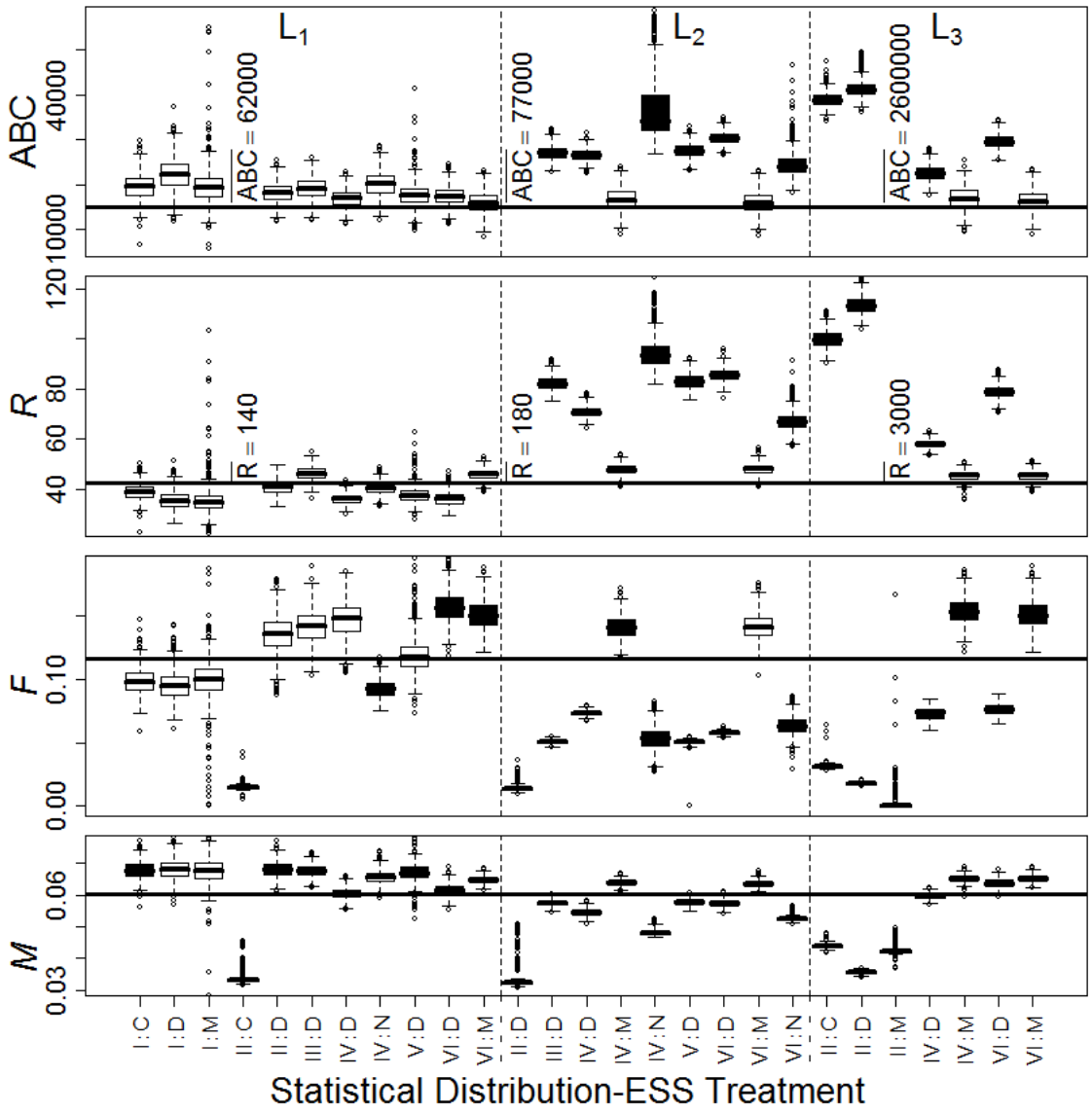

Figure 2.6. ESS_{est} boxplots for data generation scenario G_2. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) across estimation model combinations (L_1: Full likelihood structure, L_2: Approximate likelihood structure, L_3: Normalized likelihood structure, C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). For results beyond the range of the *y*-axis the mean is shown in text.

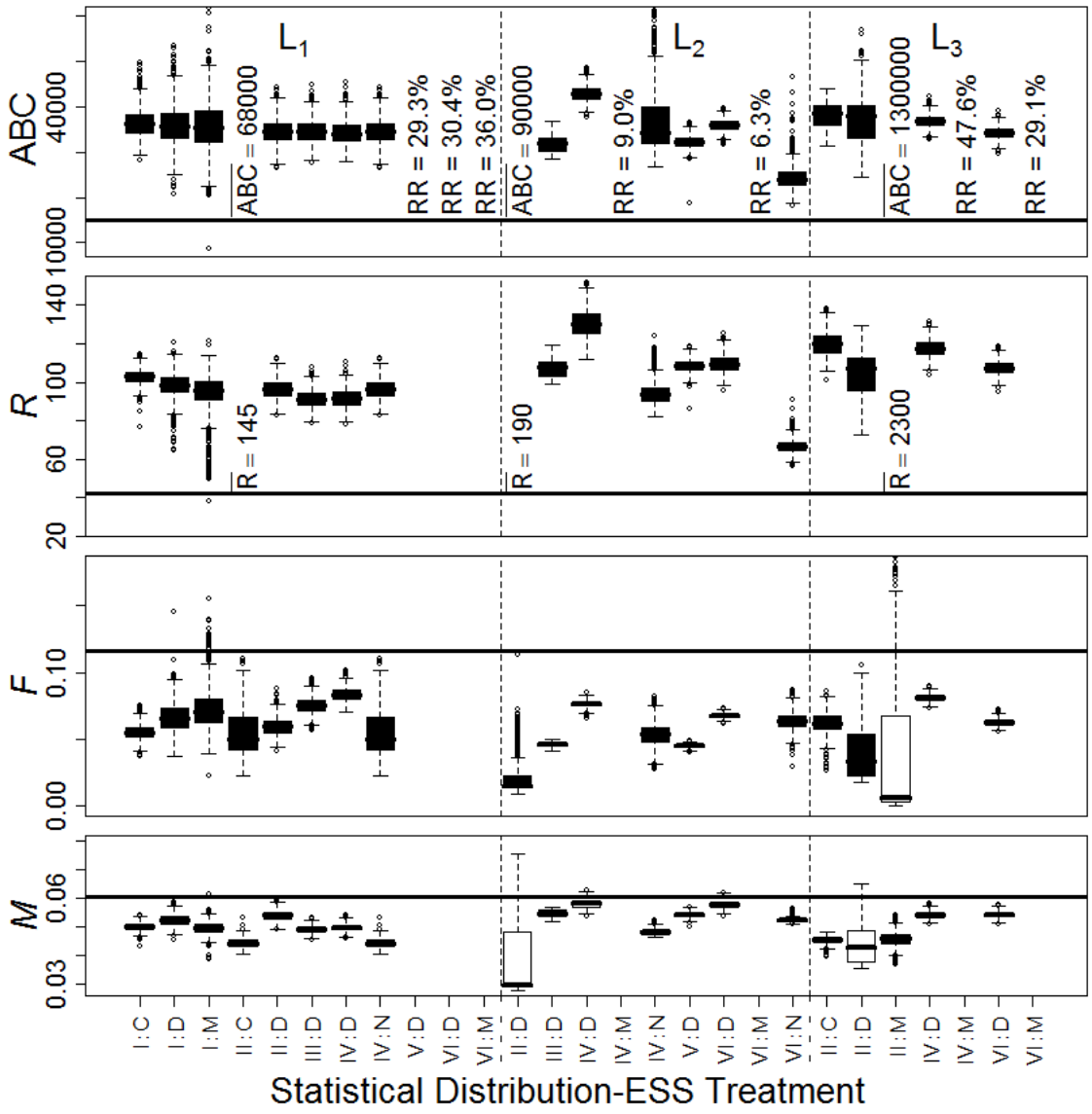

Figure 2.7. ESS_{est} boxplots for data generation scenario G_3. Estimated Acceptable Biological Catch (ABC), mean recruitment (*R*), mean fishing mortality (*F*), and natural mortality (*M*) across estimation model combinations (L_1: Full likelihood structure, L_2: Approximate likelihood structure, L_3: Normalized likelihood structure, C: compound multinomial-Dirichlet, D: Dirichlet, M: multinomial, N: normal). For results beyond the range of the *y*-axis the mean is shown in text.

Appendix 2.A: Likelihood structure derivations

2.A.1 Introduction

In this appendix the negative log-likelihood function for three types of likelihood structures are derived for the normal, multinomial, Dirichlet, and compound multinomial-Dirichlet distributions: (1) the full likelihood structure, (2) the approximate likelihood structure, and (3) the normalized likelihood structure. The full likelihood does not omit any terms in the probability density function, the approximate likelihood omits constant terms, and the normalized likelihood is determined by dividing by the likelihood value when the model predicted equals the observed value (i.e., replace the predicted proportion with the observed proportion).

2.A.2 Notation

y_0 – Starting year (1961 for Pacific Ocean Perch)

Y – Total number of years (47)

a_0 – Beginning age (age-2)

A – Total number of ages (24)

$\sigma_{a,y}$ – Standard deviation at age-a in year-y

$p_{a,y}$ – Observed proportion at age-a in year-y

$\hat{p}_{a,y}$ - Predicted proportion at age-a in year-y

n_y – Sample size in year-y (ESS when implementing likelihood)

2.A.3 Normal distribution

Full likelihood structure:

$$L = \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\frac{1}{\sigma_{a,y}\sqrt{2\pi}}e^{-\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{1}{2}\ln 2\pi + \ln\sigma_{a,y} + \frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}\right]$$

$$= \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{1}{2}\left(\ln 2\pi + \ln\left[p_{a,y}\left(1-p_{a,y}\right)\right] - \ln n_y\right) + n_y\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2p_{a,y}\left(1-p_{a,y}\right)}\right]$$

Approximate likelihood structure:

$$L = \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\frac{1}{\sigma_{a,y}\sqrt{2\pi}}e^{-\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}} \propto \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}e^{-\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}$$

$$= \sum_{y=y_0}^{Y}\frac{n_y}{2}\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{p_{a,y}\left(1-p_{a,y}\right)}$$

Normalized likelihood structure:

$$L = \frac{L_{0,N}}{L_{obs}} = \frac{\prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\frac{1}{\sigma_{a,y}\sqrt{2\pi}}e^{-\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}}}{\prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\frac{1}{\sigma_{a,y}\sqrt{2\pi}}e^{-\frac{\left(p_{a,y}-p_{a,y}\right)^2}{2\sigma_{a,y}^2}}} = \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}e^{-\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2\sigma_{a,y}^2}$$

$$= \sum_{y=y_0}^{Y}\frac{n_y}{2}\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{p_{a,y}\left(1-p_{a,y}\right)}$$

2.A.4 Multinomial distribution

Full likelihood structure:

Let $x_{a,y} = n_y p_{a,y}$

$$L = \prod_{y=y_0}^{Y}\left[\frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,x_{a,y}}\right] = \prod_{y=y_0}^{Y}\left[\frac{n_y\Gamma(n_y)}{\prod_{a=a_0}^{A} x_{a,y}\Gamma(x_{a,y})}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,x_{a,y}}\right]$$

$$= \prod_{y=y_0}^{Y}\left[\frac{n_y\Gamma(n_y)}{\prod_{a=a_0}^{A} n_y p_{a,y}\Gamma(n_y p_{a,y})}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,n_y p_{a,y}}\right]$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y}\left[(A-1)\ln n_y - \ln\Gamma(n_y) + \sum_{a=a_0}^{A}\left[\ln p_{a,y} + \ln\Gamma(n_y p_{a,y}) - n_y p_{a,y}\ln\hat{p}_{a,y}\right]\right]$$

Approximate likelihood structure:

$$L = \prod_{y=y_0}^{Y}\left[\frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,x_{a,y}}\right] \propto \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,x_{a,y}} \Rightarrow -\ln L = -\sum_{y=y_0}^{Y} n_y \sum_{a=a_0}^{A} p_{a,y}\ln\hat{p}_{a,y}$$

Normalized likelihood structure:

$$L = \frac{L}{L_{obs}} = \frac{\prod_{y=y_0}^{Y}\left[\frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!}\prod_{a=a_0}^{A}\hat{p}_{a,y}^{\,x_{a,y}}\right]}{\prod_{y=y_0}^{Y}\left[\frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!}\prod_{a=a_0}^{A} p_{a,y}^{\,x_{a,y}}\right]} = \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\left(\frac{\hat{p}_{a,y}}{p_{a,y}}\right)^{x_{a,y}} = \prod_{y=y_0}^{Y}\prod_{a=a_0}^{A}\left(\frac{\hat{p}_{a,y}}{p_{a,y}}\right)^{n_y p_{a,y}}$$

$$\Rightarrow -\ln L = -\sum_{y=y_0}^{Y} n_y \sum_{a=a_0}^{A} p_{a,y}\left(\ln\hat{p}_{a,y} - \ln p_{a,y}\right)$$

2.A.5 Dirichlet distribution

Full likelihood structure:

$$\text{Let } \alpha_{a,y} = n_y \hat{p}_{a,y},\ x_{a,y} = n_y p_{a,y}$$

$$L = \prod_{y=y_0}^{Y} \frac{\Gamma(n_y)}{\prod_{a=a_0}^{A} \Gamma(\alpha_{a,y})} \prod_{a=a_0}^{A} p_{a,y}^{\alpha_{a,y}-1} = \prod_{y=y_0}^{Y} \frac{\Gamma(n_y)}{\prod_{a=a_0}^{A} \Gamma(n_y \hat{p}_{a,y})} \prod_{a=a_0}^{A} p_{a,y}^{n_y \hat{p}_{a,y}-1}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y} \left[\sum_{a=a_0}^{A} \left[\ln \Gamma(n_y \hat{p}_{a,y}) - (n_y \hat{p}_{a,y} - 1) \ln p_{a,y} \right] - \ln \Gamma(n_y) \right]$$

Approximate likelihood structure:

$$L = \prod_{y=y_0}^{Y} \frac{\Gamma(n_y)}{\prod_{a=a_0}^{A} \Gamma(\alpha_{a,y})} \prod_{a=a_0}^{A} p_{a,y}^{\alpha_{a,y}-1} \propto \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{p_{a,y}^{\alpha_{a,y}}}{\Gamma(\alpha_{a,y})} = \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{p_{a,y}^{n_y \hat{p}_{a,y}}}{\Gamma(n_y \hat{p}_{a,y})}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y} \sum_{a=a_0}^{A} \left[\ln \Gamma(n_y \hat{p}_{a,y}) - n_y \hat{p}_{a,y} \ln p_{a,y} \right]$$

Normalized likelihood structure:

$$L = \frac{L}{L_{obs}} = \frac{\prod_{y=y_0}^{Y} \frac{\Gamma(n_y)}{\prod_{a=a_0}^{A} \Gamma(\alpha_{a,y})} \prod_{a=a_0}^{A} p_{a,y}^{\alpha_{a,y}-1}}{\prod_{y=y_0}^{Y} \frac{\Gamma(n_y)}{\prod_{a=a_0}^{A} \Gamma(x_{a,y})} \prod_{a=a_0}^{A} p_{a,y}^{x_{a,y}-1}} = \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{\Gamma(x_{a,y})}{\Gamma(\alpha_{a,y})} p_{a,y}^{\alpha_{a,y}-x_{a,y}}$$

$$= \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{\Gamma(n_y p_{a,y})}{\Gamma(n_y \hat{p}_{a,y})} p_{a,y}^{n_y \hat{p}_{a,y} - n_y p_{a,y}}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y} \left[\sum_{a=a_0}^{A} \left[\ln \Gamma(n_y \hat{p}_{a,y}) - \ln \Gamma(n_y p_{a,y}) + n_y (p_{a,y} - \hat{p}_{a,y}) \ln p_{a,y} \right] \right]$$

2.A.6 Compound multinomial-Dirichlet distribution

Full likelihood structure:

$$\text{Let } \alpha_{a,y} = n_y \hat{p}_{a,y},\ x_{a,y} = n_y p_{a,y}$$

$$L = \prod_{y=y_0}^{Y} \frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!} \frac{\Gamma(n_y)}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{\Gamma(\alpha_{a,y})} = \prod_{y=y_0}^{Y} \frac{\Gamma(n_y + 1)}{\prod_{a=a_0}^{A} \Gamma(x_{a,y} + 1)} \frac{\Gamma(n_y)}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{\Gamma(\alpha_{a,y})}$$

$$= \prod_{y=y_0}^{Y} \frac{n_y \Gamma(n_y)^2}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{x_{a,y}\Gamma(\alpha_{a,y})\Gamma(x_{a,y})} = \prod_{y=y_0}^{Y} \frac{n_y \Gamma(n_y)^2}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(n_y \hat{p}_{a,y_{a,y}} + n_y p_{a,y})}{n_y p_{a,y} \Gamma(n_y \hat{p}_{a,y_{a,y}})\Gamma(n_y p_{a,y})}$$

$$\Rightarrow -\ln L = -\sum_{y=y_0}^{Y} \left[(1-A)\ln n_y + 2\ln\Gamma(n_y) - \ln\Gamma(2n_y) - \sum_{a=a_0}^{A} \left[\begin{matrix} \ln\Gamma(n_y(\hat{p}_{a,y} + p_{a,y})) - \ln p_{a,y} \\ -\ln\Gamma(n_y \hat{p}_{a,y}) - \ln\Gamma(n_y p_{a,y}) \end{matrix} \right] \right]$$

Approximate likelihood structure:

$$L = \prod_{y=y_0}^{Y} \frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!} \frac{\Gamma(n_y)}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{\Gamma(\alpha_{a,y})} \propto \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{\Gamma(\alpha_{a,y})}$$

$$= \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{\Gamma(n_y \hat{p}_{a,y} + n_y p_{a,y})}{\Gamma(n_y \hat{p}_{a,y})} \Rightarrow -\ln L = \sum_{y=y_0}^{Y} \sum_{a=a_0}^{A} \left[\ln\Gamma(n_y \hat{p}_{a,y}) - \ln\Gamma(n_y(\hat{p}_{a,y} + p_{a,y})) \right]$$

Normalized likelihood structure:

$$L = \frac{L}{L_{obs}} = \frac{\prod_{y=y_0}^{Y} \frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!} \frac{\Gamma(n_y)}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(\alpha_{a,y} + x_{a,y})}{\Gamma(\alpha_{a,y})}}{\prod_{y=y_0}^{Y} \frac{n_y!}{\prod_{a=a_0}^{A} x_{a,y}!} \frac{\Gamma(n_y)}{\Gamma(2n_y)} \prod_{a=a_0}^{A} \frac{\Gamma(2x_{a,y})}{\Gamma(x_{a,y})}} = \prod_{y=y_0}^{Y} \prod_{a=a_0}^{A} \frac{\Gamma(n_y p_{a,y})\Gamma(n_y \hat{p}_{a,y} + n_y p_{a,y})}{\Gamma(n_y \hat{p}_{a,y})\Gamma(2n_y p_{a,y})}$$

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y} \sum_{a=a_0}^{A} \left[\ln\Gamma(n_y \hat{p}_{a,y}) - \ln\Gamma(n_y p_{a,y}) + \ln\Gamma(2n_y p_{a,y}) - \ln\Gamma(n_y(\hat{p}_{a,y} + p_{a,y})) \right]$$

Appendix 2.B: Effective sample size maximum likelihood estimate derivations

2.B.1 Introduction

In this appendix the maximum likelihood estimates (MLE) for effective sample size (ESS) in the normal, multinomial, Dirichlet, and compound multinomial-Dirichlet distributions are derived. The methods used here follow those shown in Deriso *et al.* (2007), in which the MLE is obtained using the full likelihood structure and by taking the derivative of the negative log-likelihood with respect to ESS, then setting equal to zero. In the case of the multinomial, Dirichlet, and compound multinomial-Dirichlet, closed form solutions to the MLE for ESS are not possible and the derived functions were integrated into the joint likelihood of the ASA model. For the normal distribution two cases are derived for ESS in which the variance is defined as the multinomial variance (e.g., Fournier and Archibald 1982), the first is for constant ESS across years, the second for annual ESS.

2.B.2 Notation

y_0 – Starting year (1961 for Pacific Ocean Perch)

Y – Total number of years (47)

a_0 – Beginning age (age-2)

A – Total number of ages (24)

$p_{a,y}$ – Observed proportion at age-a in year-y

$\hat{p}_{a,y}$ - Predicted proportion at age-a in year-y

n_y – Sample size in year-y (ESS when implementing likelihood)

$\Gamma(\cdot)$ – Gamma function

$\psi(\cdot)$ – Digamma function, approximated with the J.M. Bernardo AS 103 algorithm:

$$\psi(x) \cong \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6}$$

2.B.3 Normal distribution

Case I: ESS constant across years, data structure by years and ages

$$-\ln L = \frac{AY}{2}\ln 2\pi + \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{1}{2}\left(\ln\left[p_{a,y}\left(1-p_{a,y}\right)\right] - \ln n\right) + n\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2p_{a,y}\left(1-p_{a,y}\right)}\right]$$

$$\Rightarrow \frac{d}{dn}\left[-\ln L\right] = \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2p_{a,y}\left(1-p_{a,y}\right)} - \frac{1}{2n}\right] \Rightarrow \hat{n} = AY\left[\sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{p_{a,y}\left(1-p_{a,y}\right)}\right]^{-1}$$

Case II: ESS varies across years, data structure by years and ages

$$-\ln L = \frac{AY}{2}\ln 2\pi + \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{1}{2}\left(\ln\left[p_{a,y}\left(1-p_{a,y}\right)\right] - \ln n_y\right) + n_y\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2p_{a,y}\left(1-p_{a,y}\right)}\right]$$

$$\Rightarrow \frac{d}{dn}\left[-\ln L\right] = \sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{2p_{a,y}\left(1-p_{a,y}\right)} - \frac{1}{2n_y}\right] \Rightarrow \hat{n}_y = A\left[\sum_{a=a_0}^{A}\frac{\left(p_{a,y}-\hat{p}_{a,y}\right)^2}{p_{a,y}\left(1-p_{a,y}\right)}\right]^{-1}$$

2.B.4 Multinomial distribution

$$-\ln L = \sum_{y=y_0}^{Y}\left[(A-1)\ln n_y - \ln\Gamma\left(n_y\right) + \sum_{a=a_0}^{A}\left[\ln p_{a,y} + \ln\Gamma\left(n_y p_{a,y}\right) - n_y p_{a,y}\ln\hat{p}_{a,y}\right]\right]$$

$$\Rightarrow \frac{d}{dn_y}\left[-\ln L\right] = \sum_{y=y_0}^{Y}\left[\frac{(A-1)}{n_y} - \psi\left(n_y\right) + \sum_{a=a_0}^{A}p_{a,y}\left[\psi\left(n_y p_{a,y}\right) - \ln\hat{p}_{a,y}\right]\right]$$

2.B.5 Dirichlet distribution

$$\Rightarrow -\ln L = \sum_{y=y_0}^{Y}\left[\sum_{a=a_0}^{A}\left[\ln\Gamma\left(n_y\,\hat{p}_{a,y}\right)-\left(n_y\,\hat{p}_{a,y}-1\right)\ln\,p_{a,y}\right]-\ln\Gamma\left(n_y\right)\right]$$

$$\Rightarrow \frac{d}{dn_y}\left[-\ln L\right] = \sum_{y=y_0}^{Y}\left[\sum_{a=a_0}^{A}\hat{p}_{a,y}\left[\psi\left(n_y\,\hat{p}_{a,y}\right)-\ln\,p_{a,y}\right]-\psi\left(n_y\right)\right]$$

2.B.6 Compound multinomial-Dirichlet distribution

$$-\ln L = -\sum_{y=y_0}^{Y}\left[\left(1-A\right)\ln n_y + 2\ln\Gamma\left(n_y\right)-\ln\Gamma\left(2n_y\right)\right]$$

$$-\sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\ln\Gamma\left(n_y\left(\hat{p}_{a,y}+p_{a,y}\right)\right)-\ln\,p_{a,y}-\ln\Gamma\left(n_y\,\hat{p}_{a,y}\right)-\ln\Gamma\left(n_y\,p_{a,y}\right)\right]$$

$$\Rightarrow \frac{d}{dn_y}\left[-\ln L\right] = -\sum_{y=y_0}^{Y}\left[\frac{\left(1-A\right)}{n_y}+2\psi\left(n_y\right)-2\psi\left(2n_y\right)\right]$$

$$-\sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\left(\hat{p}_{a,y}+p_{a,y}\right)\ln\Gamma\left(n_y\left(\hat{p}_{a,y}+p_{a,y}\right)\right)-\hat{p}_{a,y}\,\ln\Gamma\left(n_y\,\hat{p}_{a,y}\right)-p_{a,y}\,\ln\Gamma\left(n_y\,p_{a,y}\right)\right]$$

$$=\sum_{y=y_0}^{Y}\left[\frac{\left(A-1\right)}{n_y}-2\left(\psi\left(n_y\right)-\psi\left(2n_y\right)\right)\right]$$

$$-\sum_{y=y_0}^{Y}\sum_{a=a_0}^{A}\left[\left(\hat{p}_{a,y}+p_{a,y}\right)\ln\Gamma\left(n_y\left(\hat{p}_{a,y}+p_{a,y}\right)\right)-\hat{p}_{a,y}\,\ln\Gamma\left(n_y\,\hat{p}_{a,y}\right)-p_{a,y}\,\ln\Gamma\left(n_y\,p_{a,y}\right)\right]$$

Chapter 3: Spatial Modeling with Integrated Age-Structured Assessment Models in a Changing Environment[1]

Abstract

Climate change forcing can affect the spatial distribution of fish populations, which would subsequently impair the accuracy of spatially aggregated age-structured assessment models. In this study spatially aggregated models were compared to spatially explicit models under different climate change scenarios. The climate change scenarios that were evaluated influenced the spatial distribution of walleye pollock (*Theragra chalcogramma*) in the Eastern Bering Sea by shifting the population northward. Overall, accurate estimates of total biomass resulted from both spatially aggregated and spatially explicit models, while the uncertainty in biomass estimates was reduced with a spatially explicit model. Spatially explicit models that estimated ontogenetic and climate change parameters gave smaller Deviance Information Criterion (DIC) values than those that set such parameters, although the range in DIC across replicated datasets could be large. Further, if the functional form of climate change effects on movement is unknown, using a random walk to estimate movement is feasible.

[1] Hulson, P.-J. F., and Quinn, T.J., II, in preparation. Spatial modeling with integrated age-structured assessment models in a changing environment. In preparation to be submitted to Canadian Journal of Fisheries and Aquatic Sciences.

3.1 Introduction

Spatial modeling of fish stocks for the purposes of fishery management is rarely performed. However, there are a few exceptions including school shark (*Galeorhinus galeus*, Punt *et al.* 2000), sablefish (*Anaplopoma fimbria*, Hanselman *et al.* 2007) and rock lobster (*Jasus edwardsii*, McGarvy *et al.* 2010). The primary limitation to implementing spatial models is the lack of spatially disaggregated datasets. Consequently, stock assessment scientists use spatially aggregated models that do not estimate movement, evaluating a fish stock across a broad geographic scale. With spatially disaggregated datasets and mark-recapture information, spatially explicit models can be developed that estimate movement between regions (Miller *et al.* 2008, Goethel *et al.* 2011). Within this study we consider a model that does not include spatial information or estimate movement to be a spatially aggregated model; a model that does estimate movement and include spatial data is a spatially explicit model.

An intermediate step in moving towards ecosystem based fisheries management (EBFM) is to improve current stock assessment models by developing spatially explicit stock assessment models (Quinn and Collie 2005, Goethel *et al.* 2011). In terms of making EBFM operational, spatially explicit models would evaluate interactions among aquatic species at finer spatial scales, as modeling at broad geographic scales can misinterpret fundamental population dynamics (Härkönen and Harding 2001). Spatially explicit stock assessment models contribute to understanding the interactions between fish and environmental influences within geographically distinct habitats (Mueter *et al.* 2006). Currently, spatially explicit stock assessment models are not commonly used in

fisheries management; studies are being devoted to examining and developing spatially explicit models (Punt *et al.* 2000, Miller *et al.* 2008, Goethel *et al.* 2011).

Research conducted by the Intergovernmental Panel on Climate Change (IPCC) has found that global average temperatures are rising (e.g., IPCC 2007), which could have an influence on the spatial distribution of fish species evaluated with fisheries stock assessment models (Mueter and Litzow 2008). Thus, the potential effects of climate change on the spatial distribution of fish populations need to be identified (Beamish *et al.* 2004). A hypothesized effect of climate change on the spatial distribution of fish species is a latitudinal shift in distribution (Cheung *et al.* 2009). A number of studies have examined the ecological implications of climate change through simulation (Hashioka and Yamanaka 2007, Munday *et al.* 2008), although the performance of fisheries stock assessment models when applied to fish populations influenced by climate change has not been investigated. Also, the effectiveness of spatially aggregated stock assessment models compared to spatially explicit stock assessment models when considering climate change needs clarification.

A spatially explicit stock assessment model for Eastern Bering Sea (EBS) walleye pollock (*Theragra chalcogramma*, Miller *et al.* 2008, Appendix A) was used in this study as it provides an excellent example of a fish species whose latitudinal range could be affected by climate change (Overpeck *et al.* 1997). Within the operating model we produce a hypothetical northward shift in the EBS pollock population resulting from climate change and through simulation we investigate the accuracy and precision of spatially aggregated and spatially explicit age-structured assessment models. We also

investigate the feasibility of estimating annual deviations in movement with a random walk and of estimating ontogenetic movement within spatially explicit stock assessment models.

3.2 Methods

In this study we perform simulation testing by first constructing an operating model with ‘known’ population structure and then input generated data from the operating model into the estimation model. Within the operating model hypothetical effects of climate change on spatial distribution are considered by shifting the population of pollock in the EBS to the north. The estimation models are subsequently fit to the generated data and evaluated by implementing different parameterizations for movement. Two structures of estimation models are considered, a spatially aggregated model that does not incorporate movement and a spatially explicit model that estimates movement.

3.2.1 Operating Model

The operating model used in this study was based on parameter estimates drawn from a spatially explicit ASA model for pollock (Miller *et al.* 2008). In this study, the parameter estimates from Appendix A were used as ‘true’ parameters in the operating model when no climate change occurred. This spatially explicit ASA model included two regions (NW and SE) of the Eastern Bering Sea and two commercial fishery seasons (A winter/spring season, B summer/fall season, Miller *et al.* 2008). Movement between regions occurred after each fishery season. The first movement, after the A season, was

movement from the spawning to summer feeding grounds; the second movement, after the B season, was movement from the summer feeding grounds back to the spawning grounds. Movement of pollock in the Eastern Bering Sea has been found to be related to age (Buckley *et al.* 2001), and in the operating model retention in the NW to SE regions was a decreasing function of age for both the A and B fishing seasons. The number of fish that stayed in the NW region after the A season fishery decreased by 80% per age class and decreased by 90% per age class after the B season fishery in the operating model. Further details, including parameters estimated and population dynamics equations, can be found in Appendix A and Miller *et al.* (2008).

Datasets used in the spatially explicit ASA model spanned the years 1977 to 2005 and included ages 3 to 10+ in the age composition. Commercial fishery data included total fishery yield for all years from 1977-2005, fishery yield from the NW region for the A season (1977-1987, 1991-1996, and 1998-2004) and B season (1977-1987, 1991-2004), and fishery yield from the SE region for the A and B seasons (both for years 1977-1987, 1991-2004). Fishery independent indices of abundance included the Bottom Trawl Survey (BTS) biomass from 1982-2004 and the Echo-Integration Trawl (EIT) survey in the years 1994, 1996-1997, 1999-2000, and 2002. Age composition data were included for each year of the regional and seasonal commercial fishery catch, BTS, and EIT. A hypothetical tagging program was implemented in the operating model, and was similar to the approach in Appendix A. One main difference between the current study and Appendix A was that the simulated tags were released for the last 10 years of the time

series rather than from 2000 to 2002 only. Simulated tag returns were modeled for the A season from 1995-2004 and for the B season from 1994-2004.

3.2.2 Climate Change Scenarios

In this study we considered two cases for the spatial distribution of pollock. There are a number of mechanisms through which climate change can affect the spatial distribution of fish including both broad and localized changes. In this study we simplify the number of mechanisms investigated by hypothesizing that the various mechanisms would result in a northward shift in the population. In the first (base case), movement was the same as in Appendix A, i.e., constant across years, and in the second (climate change case), movement to the NW region was increased at the end of the time series so that a northward shift in the pollock population resulted (Mueter and Litzow 2008). A number of processes in fish populations can be affected by climate change (Perry *et al.* 2005) but in this study only climate change effects on movement, rather than recruitment or natural mortality, for example, were considered.

In the operating model for the climate change case the shift to the NW region was produced by altering movement and keeping all other parameters the same as in the base case. In the climate change case movement and retention in the NW region was increased with a linear function from 1997 to 2005. The linear function used in the climate change cases was:

$$\theta_{R,S,y+1} = \theta_{R,S,y} \cdot \kappa \tag{3.1}$$

where $\theta_{R,S,y}$ was the logit transformed movement parameter that was related to the proportion of fish that stayed in region *R*, after season *S*, in year *y*, and κ was the climate change parameter. For the base case, κ was 1; thus, there was constant movement across years. To provide sufficient contrast between the base and climate change case, κ was set at 1.2 in the climate change case, which produced increased movement and retention in the NW region when applied to the movement parameters. Ontogenetic movement was also a function of κ in the climate change case following equation 3.1, resulting in time-dependent ontogenetic movement. Overall, the total biomass was similar between the base case and climate change case and the increase in biomass in the NW region was accompanied by a related decrease in the SE region (Figure 3.1).

3.2.3 Data Generation

Error in catch and index data for the commercial fishery, BTS, and EIT was generated with the lognormal distribution; the dataset predictions from Hulson *et al.* (in press) were treated as 'true' median values in the lognormal distribution. The standard deviations (SD's) used to generate the fishery yield and survey biomasses were the same SD's used in Appendix A. Commercial fishery yield in the NW and SE regions for the A and B fishing seasons were generated with a SD of 0.05. Regional BTS and EIT biomass datasets were generated with a SD of 0.2. Spatially aggregated index data were generated by summing the regional index data (i.e., sum of lognormal random variables); the SD of the non-spatial index data was actually smaller than the spatially explicit data and was

0.025 for the total fishery yield, 0.04 for the seasonal fishery yield, and 0.14 for the BTS and EIT biomass, as calculated with the Fenton and Wilkinson approximation method (Gao *et al.* 2009).

Age composition datasets for the commercial fishery, BTS, and EIT were generated with the multinomial distribution across the climate change cases. For each year of the regional and seasonal fishery age composition, regional BTS age composition, and regional EIT age composition, multinomial error was generated by using the age composition from the operating model for each dataset and a sample size of 200. The generating sample size used was smaller than the actual sample sizes collected for age composition in each case and was selected to represent potential sources of unknown error in the age composition (e.g., McAllister and Ianelli 1997). The spatially aggregated age composition was generated by simply summing the random multinomial values for the generated regional age compositions. Note that the variance of the spatially aggregated age composition was actually smaller than the variance of the spatially explicit age composition due to the increase in sample size.

Tagging data were generated in this study with an interactive multinomial – overdispersed Poisson scheme. Following previous research into the feasibility of tagging EBS pollock (Miller 2007), within the simulation 10,000 tags were released during the BTS survey. Tag releases by age were generated with the multinomial distribution based on the regional BTS age composition rather than uniformly tagging across ages as was done in Appendix A. For simplicity, the number of tags by age that were released was

assumed known and used in the spatially explicit model as annual additions to the tagged population.

Tag returns occurred during the fishing seasons, and were generated with an overdispersed Poisson distribution (Fukushima *et al.* 1998). The overdispersion parameter used for the tag return data generation was set at 2 for (*i*) the SE region during the A season, and (*ii*) the NW and SE regions during the B season. To avoid having too much error when few tags were recaptured, the overdispersion parameter was set at 1.5 for the A season in the NW region. The annual and regional number of fish with tags that were caught in the fishery were subtracted from the total number of tags in the population. The proportion of total catch that was examined for tags was set at 7% for the NW region and 39% for the SE region and was based on the proportion of catch that was landed to shoreside processors (Miller 2007). Thus, the generated total number of tags caught in the commercial fishery was multiplied by the proportion of catch delivered to shoreside processors to produce the tag return datasets used in the estimation model scenarios.

For each climate change case, catch, population index, age composition, and tag return datasets were generated 1,000 times (Efron and Tibshirani 1993) using R (R Development Core Team 2008). Note that the estimation model scenarios were fitted to the same generated datasets for each climate change case to allow for more direct comparison. For each climate change case in the operating model, error was generated in datasets fit by the spatially aggregated and spatially explicit stock assessment models and subsequently applied to the estimation model scenarios.

3.2.3 Estimation Model Scenarios

We evaluated two primary structures of ASA models in this study: a spatially aggregated ASA model (model scenario M0) and a spatially explicit ASA model (model scenarios M1-M4, Table 3.1). The spatially explicit ASA model integrated tagging data and was based on the Integrated Tagging and Catch-at-Age Analysis **(**ITCAAN) model after Maunder (2001). From this point forward the spatially aggregated ASA model will simply be referred to as the ASA model, while the spatially explicit ASA model will be referred to as the ITCAAN model. The ITCAAN model structure was the same as in Appendix A. The ASA model did not estimate movement and was similar in structure to the current ASA model used to manage pollock (Ianelli *et al.* 2007). For simplicity, the SD's used in data generation were assumed known and used in the lognormal likelihood functions in the estimation model scenarios when fitting to the generated index data for both the ASA and ITCAAN models. The estimation model scenarios were fitted to generated age composition with the Dirichlet distribution, in which the effective sample size was parameterized and estimated as a mean parameter across years for each age composition dataset (Chapter 2) for both the ASA and ITCAAN models. Tag returns were fitted with the Poisson distribution.

Parameterization in the ITCAAN model for model scenarios M1-M4 focused on parameters for ontogenetic (γ) and climate change (κ) effects on movement; the ontogenetic and climate change parameters were misspecified, known, or estimated (Table 3.1). The ontogenetic and climate change parameters were misspecified by setting

the parameter value to 1; i.e., constant movement was assumed across ages (for the ontogenetic parameters) and years (for the climate change parameter). Estimation model scenario M1 was the worst case scenario, in which both the ontogenetic and climate change parameters were misspecified. In estimation model scenario M2, both the ontogenetic and climate change parameters were known, and in M3 they were estimated.

Model scenario M4 was built to reflect the situation in which the functional form of climate change was unknown and regional and seasonal movement was estimated with a random walk (Wilberg and Bence 2006). If a random walk was applied to each regional and seasonal movement parameter and ontogenetic parameter the number of parameters estimated would increase by 6 times the number of years modeled (an increase of 168 parameters). To reduce the number of parameters and the computational time required to run such complicated parameterizations, we assumed that the relative annual changes in the movement and ontogenetic parameters would have a similar relationship with climate change forcing. Thus, we applied a single time series of annual deviations to the initial regional and seasonal movement parameters and ontogenetic parameters, resulting in an increase in the number of parameters estimated of only 28.

3.2.4 Model Comparison

Model comparison among the estimation model scenarios focused on bias, particularly in the estimated biomass and in other parameters that were comparable between the ASA and ITCAAN models. The ASA model estimated 98 parameters while the ITCAAN models' number of parameters ranged from 193 to 223. To reduce the

number of results shown, comparison of model scenarios focused on biomass estimated in the last year (both total and regional, B_Y, $B_{NW,Y}$, $B_{SE,Y}$), mean recruitment ($\overline{R}$) and BTS catchability (q_{BTS}). We focus on bias in total or regional biomass in the last year of the model because accurate biomass in the last year is critical to management quantities. We evaluated mean recruitment and BTS catchability because these parameters were estimated in both the ASA and ITCAAN models. We also investigated the accuracy and precision of ontogenetic movement for ITCAAN model scenarios in which it was estimated. For the results, boxplots are shown, where the box represents the inter-quartile range from the estimation models and the whiskers represent the 95% confidence interval. We consider an estimate to be significantly biased if the value from the operating model is outside the 95% confidence intervals.

The Deviance Information Criterion (DIC, Spiegelhalter *et al.* 2002, Wilberg and Bence 2008) was also used to compare the ITCAAN estimation model scenarios. The DIC value was computed for each estimation model scenario across 100 randomly selected dataset replicates. The number of replicates examined was smaller than the total number of replicates (1,000) due to the extensive amount of time required to run the Markov-Chain Monte Carlo (MCMC) simulation. Across the 100 replicated datasets, samples from MCMC chains were stored for each estimation model scenario (e.g., Gelman *et al.* 2004). The MCMC chain length was 500,000 for each replicated dataset, and to ensure chain convergence and reduce autocorrelation, 250 samples of the MCMC chain were selected for DIC calculation (details provided in Appendix 3.A). DIC was computed as (following Spiegelhalter *et al.* 2002):

$$DIC = \overline{D(\Theta)} + p_D \tag{3.2}$$

where $D(\cdot)$, the 'Bayesian deviance' was defined as:

$$D(\Theta) = -2 \ln L(y|\Theta) + 2 \ln L(y) \tag{3.3}$$

in which $L(y|\Theta)$ is the likelihood function for the data, *y*, given the estimated parameters, Θ. The effective number of parameters, p_D, was calculated as:

$$p_D = \overline{D(\Theta)} - D(\overline{\Theta}) \tag{3.4}$$

where $\overline{D(\Theta)}$ was the average deviance across the MCMC samples of parameter estimates, and $D(\overline{\Theta})$ was the deviance at the average of the parameter estimates across the MCMC samples. Computing the DIC and p_D value across the 100 replicated datasets allowed comparison of the range in DIC among estimation model scenarios. The MCMC simulation was performed for the 100 replicates in ADMB through code that was linked to data generation in R.

3.3 Results

All of the model scenarios resulted in accurate estimation of total biomass in the last year for the base case. Significant bias emerged from model scenario M1 when the population shifted northward in the climate change case (top panel, Figure 3.2). For both the base case and climate change case, the largest variability in total biomass resulted from estimation model scenario M0 (top panel, Figure 3.2). For model scenario M1, in the base case accurate estimates were obtained for $\overline{R}$ and q_{BTS}, but significant bias resulted under the climate change case (middle and bottom panels, Figure 3.2). The uncertainties in $\overline{R}$ and q_{BTS} were comparable between model scenario M0 and M1 – M4. Bias in total biomass under the climate change case for model scenario M1 was due to significant bias in biomass in both regions (Figure 3.3). Model scenarios M2 – M4 resulted in accurate regional biomass for both the base case and climate change case (Figure 3.3).

For model scenarios M3 and M4, the ontogenetic parameters estimated were accurate under both the base case and climate change case (Figure 3.4). For the NW region ontogenetic parameter after the A season, uncertainty increased in the base case at the end of the time series for both model scenarios M3 and M4, while uncertainty in the ontogenetic parameter after the B season was similar across years (Figure 3.4). Both model scenarios M3 and M4 accurately estimated the shift in the ontogenetic parameters under the climate change case (Figure 3.4).

In the base case the DIC values for estimation model scenarios M1 – M4 were not significantly different (Figure 3.5). For both the base case and climate change case the

preferred model scenario was M3, followed by M2 and then M4. In the climate change case the DIC values for model scenario M1 were significantly larger than the DIC values from model scenarios M2-M4 (Figure 3.5). While model scenario M3 was preferred for both the base case and climate change case, the range in DIC values was similar among model scenarios M2-M4 for both the base case and climate change case. For each model scenario, the effective number of parameters across replicates was smaller than the actual number of parameters estimated (Figure 3.5). The trends in effective number of parameters were similar for model scenarios M2-M4; for model scenario M1 the range in the effective number of parameters was small in the base case and large in the climate change case (Figure 3.5).

3.4 Discussion

Accurate estimation of total biomass was obtained with both the spatially aggregated and spatially explicit stock assessment models under the hypothetical climate change case investigated in this study. However, the uncertainty in biomass was larger with the spatially aggregated than with the spatially explicit model. This result is somewhat perplexing because the uncertainty generated in the datasets fit by the spatially aggregated model was smaller than the uncertainty generated in the datasets fit by the spatially explicit stock assessment model due to increases in annual sample size. Therefore, the decrease in estimated variance in total biomass in the spatially explicit stock assessment model must be due to the additional information that tagging data provided in the ITCAAN model compared to the ASA model. While accurate estimation

of total biomass is possible with a spatially aggregated stock assessment model, the reduced uncertainty in a spatially explicit stock assessment model would better inform management decisions on size and allocation of catch quotas, even if the spatial distribution was affected by climate change.

When bias in estimates of regional biomass was present in the spatially explicit model, it occurred when the ontogenetic and climate change parameters were misspecified; that is, when there were alterations in the spatial distribution because of climate change that were not properly accounted for. The model scenario (M1) that resulted in biased estimates of regional biomass underestimated the biomass in the NW region and overestimated the biomass in the SE region, because the model was unable to accurately estimate the northward shift in the population. In a real-world scenario this bias could be identified through the fit to regional index data (i.e., catch or survey biomass), as the fit to each region's index data degraded as climate change forcing shifted the population northward. In terms of management, suboptimal spatial allocation of catch would result in overexploitation in one region, and under-exploitation in the other. If ontogenetic effects are unknown, bias in stock assessment and incorrect management actions could result (Lindeman *et al.* 2000). We extend this concept by showing that the same is true for climate change forcing. We have shown that if climate change effects on ontogenetic movement are not accounted for, a spatially explicit stock assessment model would potentially have no improvement over a spatially aggregated model, the consequence being suboptimal allocation of fishing effort.

Within the spatially explicit ITCAAN model, it was not necessary to specify the ontogenetic and climate change parameters rather, it was possible to estimate them. The main caveat to this, though, is that the functional forms of the ontogenetic and climate change effects were known, thus allowing for correct parameterization. In reality, the functional form may not be known, creating possible sources of model specification error. We did find that if the functional form of climate change was unknown, a random walk (Wilberg and Bence 2006) could be used instead to estimate annual movement and result in unbiased estimation of regional biomass.

Among the ITCAAN estimation model scenarios, the average DIC value from the random walk model scenario (M4) was not significantly different than the "preferred" model (M3) for both the base case and climate change case. The effect of model specification error for ontogenetic effects remains unclear, because evaluation of misspecifying the functional form of ontogenetic movement was beyond the scope of this study. While it was possible to accurately estimate ontogenetic parameters, they can become confounded with movement parameters, something that must be kept in mind when parameterizing ontogenetic movement. Further, parameterizing age-specific movement can quickly increase the number of parameters estimated, as ontogenetic movement in real world circumstances could vary by age and over time.

Evaluating DIC across replicated datasets showed that there can be a large range in DIC values for a given estimation model scenario and that there weren't significant differences in DIC between the ITCAAN estimation model scenarios under the base case. Investigating the differences in DIC, however, showed that the estimation model scenario

that estimated both the ontogenetic and climate change parameters was preferred. A drawback in this study was the inability to directly compare the spatially aggregated to the spatially explicit estimation model scenarios using an objective criterion like DIC. The reason for this was the structure of the joint likelihoods used in the two model structures and the data to which the models were fit. Using more complex likelihood functions such as the Dirichlet distribution presents difficulties when comparing between model fits to spatially aggregated and spatially explicit data sources. In future developments of model comparison criteria we find that two factors should be considered: (1) developing model comparison criteria that produce more precise results across replicated datasets; and (2) developing model comparison criteria that can compare between different spatial structures of models that inherently use the same samples obtained from either fishery or fishery-independent sources.

Both the spatially aggregated and spatially explicit stock assessment models have drawbacks and advantages (e.g., Quinn and Collie 2005). A primary drawback of the spatially aggregated model is the inability to estimate regional biomass. In terms of management, it could become difficult to properly manage fishing effort with a spatially aggregated model if climate change were to induce a northward shift in fish populations (Cheung *et al.* 2009). An advantage of a spatially aggregated model is that accurate estimation of total biomass can be obtained without having to know ontogenetic or climate change parameters. A drawback of the spatially explicit model is potential model specification error when considering ontogenetic (Elsdon and Gillanders 2003) or climate change effects on movement. Advantages of the spatially explicit model over a spatially

aggregated model are the ability to estimate regional biomass and increased precision in total biomass, which would have implications to management because recent work has been devoted to incorporating uncertainty in model predictions when setting harvest recommendations (NRC 1998). We have also shown that a spatially explicit model provides the ability to identify potential effects on fish population dynamics by climate change and adjust management of commercial fisheries based on changes in the spatial distribution of fish species. As an intermediate step to EBFM we recommend that spatially explicit models continue to be developed and that tagging programs be implemented so that changes in fish spatial distribution due to climate change can be identified.

3.5 Acknowledgments

This publication is the result of research sponsored by Alaska Sea Grant with funds from the National Oceanic and Atmospheric Administration Office of Sea Grant, Department of Commerce, under grant no. NA06OAR4170013 (project number R/31-16) and grant no. NA10OAR4170097 (project number R/31-19), and from the University of Alaska with funds appropriated by the state and the Cooperative Institute for Arctic Research/Alaska Fisheries Science Center (CIFAR/AFSC) fellowship program. We would also like to thank Dr.'s Milo Adkison, Brenda Norcross, Dana Hanselman, and Gary Marty for helpful comments and advice.

3.6 References

Beamish, R. J., Mahken, C., and Neville, C. M. 2004. Evidence that reduced early marine growth is associated with lower marine survival of Coho salmon. Transactions of the American Fisheries Society, 133: 26–33.

Buckley, T. W., Greig, A., Ianelli, J., Livingston, P., and Walters, G. 2001. Spatial distribution and ontogenetic movement of pollock in the eastern Bering Sea. Poster presented to PICES, Alaska Fisheries Science Center, Seattle, WA.

Cheung, W. L., Lam, W. Y., Sarmiento, J. L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish and Fisheries, 10: 235–251.

Efron, B., and Tibshirani, R. J. 1993. An introduction to the Bootstrap. Chapman and Hall, New York.

Elsdon, T. S., and Gillanders, B. M. 2003. Reconstructing migratory patterns of fish based on environmental influences on otoliths chemistry. Reviews in Fish Biology and Fisheries, 13: 219–235.

Fournier, D. 1996. An introduction to AD Model Builder for use in nonlinear modeling and statistics. Nanaimo. Nanaimo, B.C., Otter Research Ltd. 56 pp.

Fukushima, M., Quinn, T. J., II, and Smoker, W. W. 1998. Estimation of egg lost from superimposed pink salmon (*Oncorhynchus gorbuscha*) redds. Canadian Journal of Fisheries and Aquatic Sciences, 55: 618–625.

Gao, X., Xu, H., and Ye, D. 2009. Asymptotic behavior of tail density for sum of correlated lognormal variables. International Journal of Mathematics and Mathematical Sciences, 2009: 1–28.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. Bayesian Data Analysis 2nd ed. Chapman and Hall/CRC, Boca Raton, Florida.

Goethel, D. R., Quinn, T. J., II, and Cadrin, S. X. 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Reviews in Fisheries Science, 19(2): 119–136.

Hanselman, D., Lunsford, C. R., Fujioka, J. T., and Rodgveller, C. J. 2007. Alaska Sablefish Assessment for 2008. *In* Stock assessment and fishery evaluation report for 2008, pp. 195-312. North Pacific Fisheries Management Council, Anchorage, AK.

Härkönen, T., and Harding, K. C. 2001. Spatial structure of harbour seal populations and the implications thereof. Canadian Journal of Zoology, 79: 2115–2127.

Hashioka, T., and Yamanaka, Y. 2007. Ecosystem change in the western North Pacific associated with global warming using 3D-NEMURO. Ecological Modeling, 202: 95–104.

Ianelli, J. N., Barbeaux, S., Honkalehto, T., Kotwicki, S., Aydin, K., and Williamson, N. 2007. Eastern Bering Sea Walleye Pollock. *In* Stock Assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands as projected for 2008. North Pacific Management Council, Anchorage, AK.

Intergovernmental Panel on Climate Change (IPCC) 2007. Climate Change 2007: The Physical Science Basis. In Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., and Miller, H.L.. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996 pp.

Lindeman, K. C., Pugliese, R., Waugh, G. T., and Ault, J. S. 2000. Developmental patterns within a multispecies reef fishery: management applications for essential fish habitats and protected areas. Bulletin of Marine Science, 66(3): 929–956.

Maunder, M. 2001. Integrated Tagging and Catch-at-Age Analysis (ITCAAN): Model development and simulation testing. *In* Spatial Processes and Management of Fish Populations, pp. 123–146. Ed. by G. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and D. Witherell. Alaska Sea Grant College Program No. AK-SG-01-02, Univeristy of Alaska, Fairbanks, AK.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54(2): 284–300.

McGarvy, R., Linnane, A. J., Feenstra, J. E., Punt, A. E., and Matthews, J. M. 2010. Integrating recapture-conditioned movement estimation into spatial stock assessment: A south Australian lobster fishery application. Fisheries Research, 105(2): 80–90.

Miller, S. 2007. Estimating movement with a spatially-explicit stock assessment model of Eastern Bering Sea pollock (*Theragra chalcogramma*). M. S. thesis, School of Fisheries and Ocean Sciences, University of Alaska, Fairbanks, AK.

Miller, S., Ianelli, J., and Quinn, T. J., II, 2008. Estimation of age-specific migration in an age-structured model. *In* Resiliency of Gadid Stocks to Fishing and Climate Change, pp. 161–178. Ed. by G. Kruse, K. Drinkwater, J. N. Ianelli, J. S. Link, D. L. Stram, V. Wespestad, and D. Woodby. Alaska Sea Grant College Program Report AK-SG-08-01, University of Alaska, Fairbanks.

Mueter, F. J., Ladd, C., Palmer, M. C., and Norcross, B. L. 2006. Bottom-up and topdown controls of walleye pollock (*Theragra chalcogramma*) on the Eastern Bering Sea shelf. Progress in Oceanography, 68: 152–183.

Mueter, F. J., and Litzow, M. A. 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecological Applications, 18(2): 309–320.

Munday, P. L., Jones, G. P., Pratchett, M. S. and Williams, A. J. 2008. Climate change and the future for coral reef fishes. Fish and Fisheries, 9: 261–285.

National Research Council (NRC). 1998. Improving Fish Stock Assessments. National Academy Press, Washington, DC.

Overpeck, J., Hughen, K., Hardy, D., Bradley, R., Case, R., Douglas, M., Finney, B., Gajewski, K., Jacoby, G., Jennings, A., Lamoureux, S., Lasca, A., MacDonald, G., Moore, J., Retelle, M., Smith, S., Wolfe, A., and Zielinski, G. 1997. Arctic environmental change of the last four centuries. Science, 278: 1251–1256.

Perry, A. L., Low, P. J., Ellis, J. R, and Reynolds, J. D. 2005. Climate change and distribution shifts in marine fishes. Science, 308: 1912–1915.

Prager, M. H., and Shertzer, K. W. 2010. Deriving Acceptable Biological Catch from the Overfishing Limit: Implications for Assessment Models. North American Journal of Fisheries Management, 30(1): 289–294.

Punt, A., Pribac, F., Walker, T. I., Taylor, B. L., and Prince, J. D. 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Marine and Freshwater Research, 51(3): 205–220.

Quinn, T. J., II, and Collie, J. S. 2005. Sustainability in single-species population models. Philosophical Transactions of the Royal Society B, 360: 147–162.

R Development Core Team 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Intergovernmental Panel on Climate Change (IPCC) 2007. Climate Change 2007: The Physical Science Basis. *In* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Ed. by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996 pp.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, 64(4): 583–639.

Wilberg, M. J., and Bence, J. R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences, 63: 2275–2285.

Wilberg, M. J., and Bence, J. R. 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. Fisheries Research, 93: 212–221.

Table 3.1. Model scenarios investigated across climate change cases. Model scenarios are defined by model structure and ontogenetic and climate change parameterizations.

Model Scenario	Model Structure	Ontogenetic parameter (γ)	Climate Change parameter (κ)
M0	ASA	NA	NA
M1	ITCAAN	Misspecified	Misspecified
M2		Known	Known
M3		Estimated	Estimated
M4		Estimated with random walk	

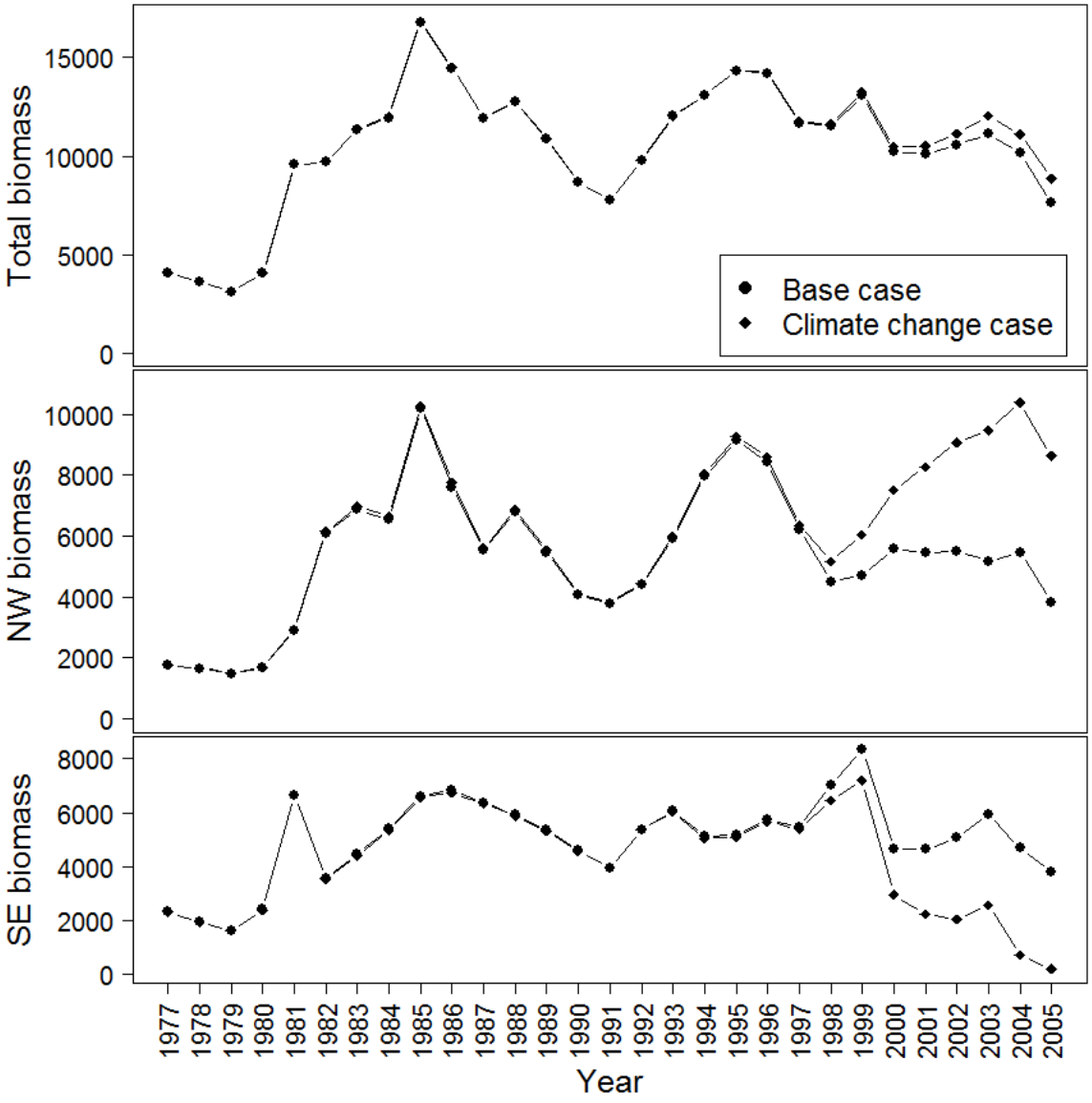

Figure 3.1. Total and regional biomass from the operating model. Total, northwest (NW), and southeast (SE) biomass (mt) from 1977-2005 for the operating model under the base case and climate change case.

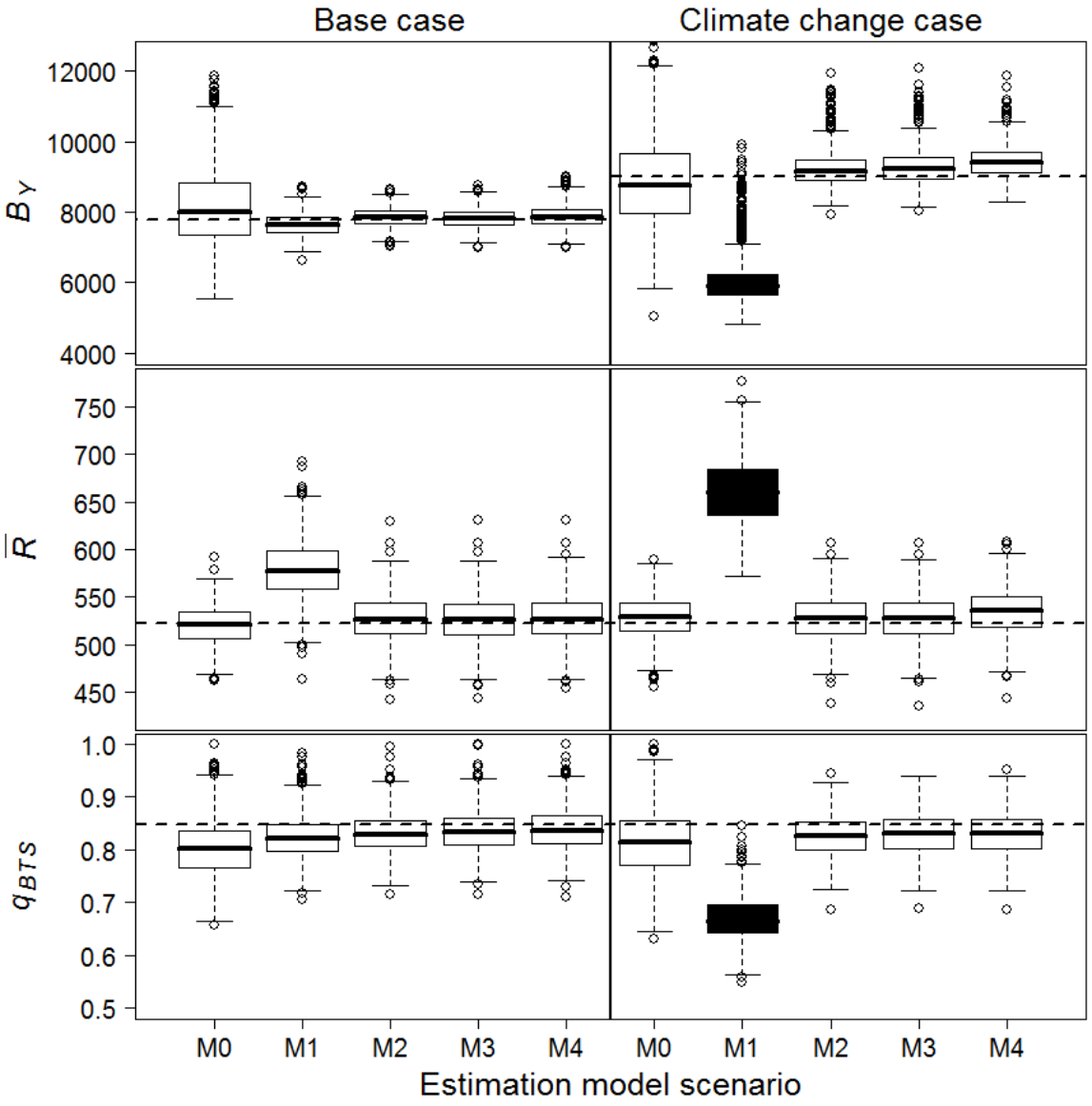

Figure 3.2. Biomass and parameter comparison among model scenarios. Total biomass estimated in the last year (B_Y), mean recruitment ($\overline{R}$), and BTS catchability (q_{BTS}) from estimation model scenarios under the base case and climate change case. The horizontal dashed lines are the values from the operating model and boxes with black fill highlight instances of significant bias.

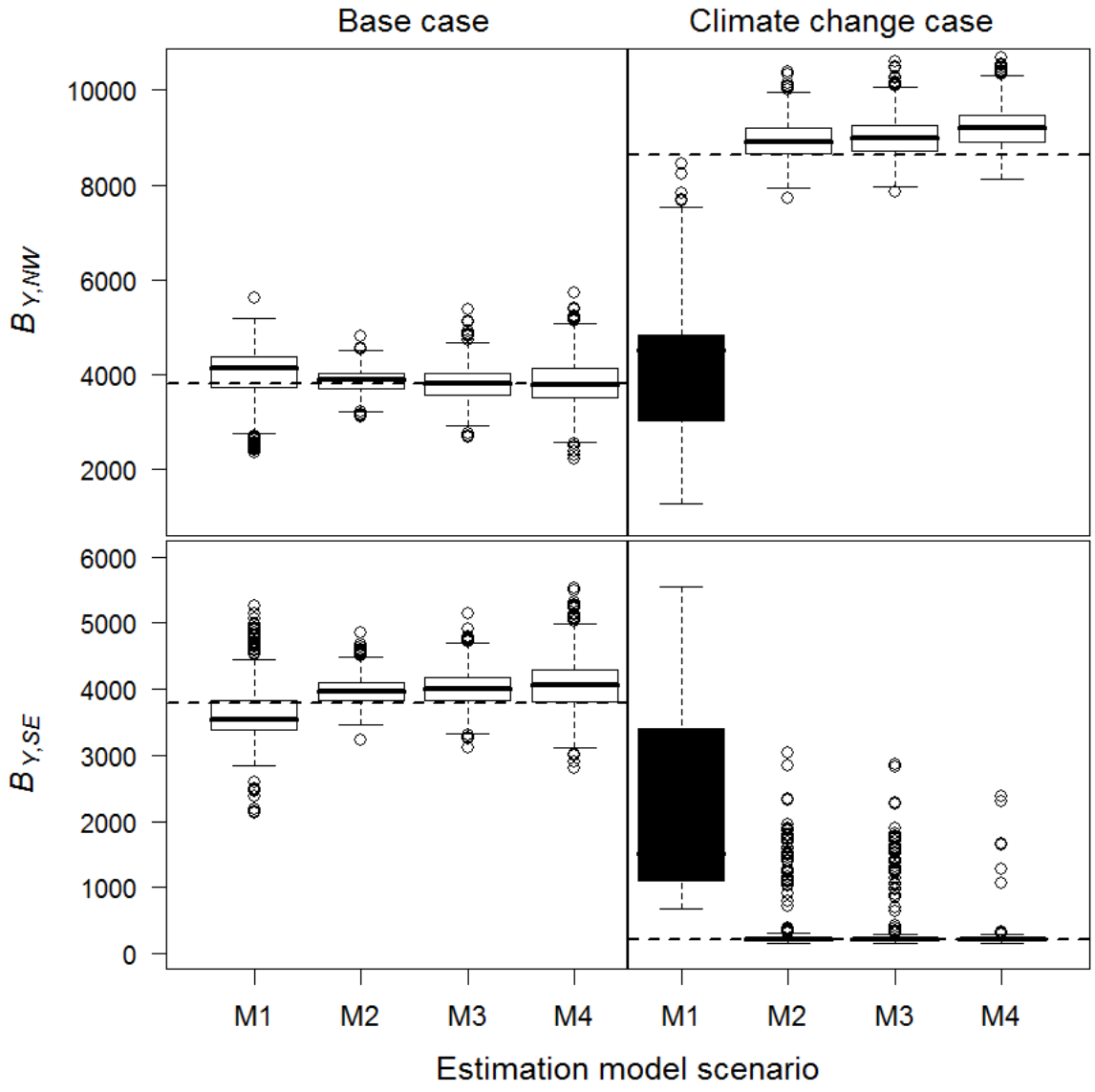

Figure 3.3. Regional biomass comparison among model scenarios. Regional biomass estimated in the last year (NW: $B_{Y,NW}$, SE: $B_{Y,SE}$) from estimation model scenarios under the base case and climate change case. The horizontal dashed lines are the values from the operating model and boxes with black fill highlight instances of significant bias.

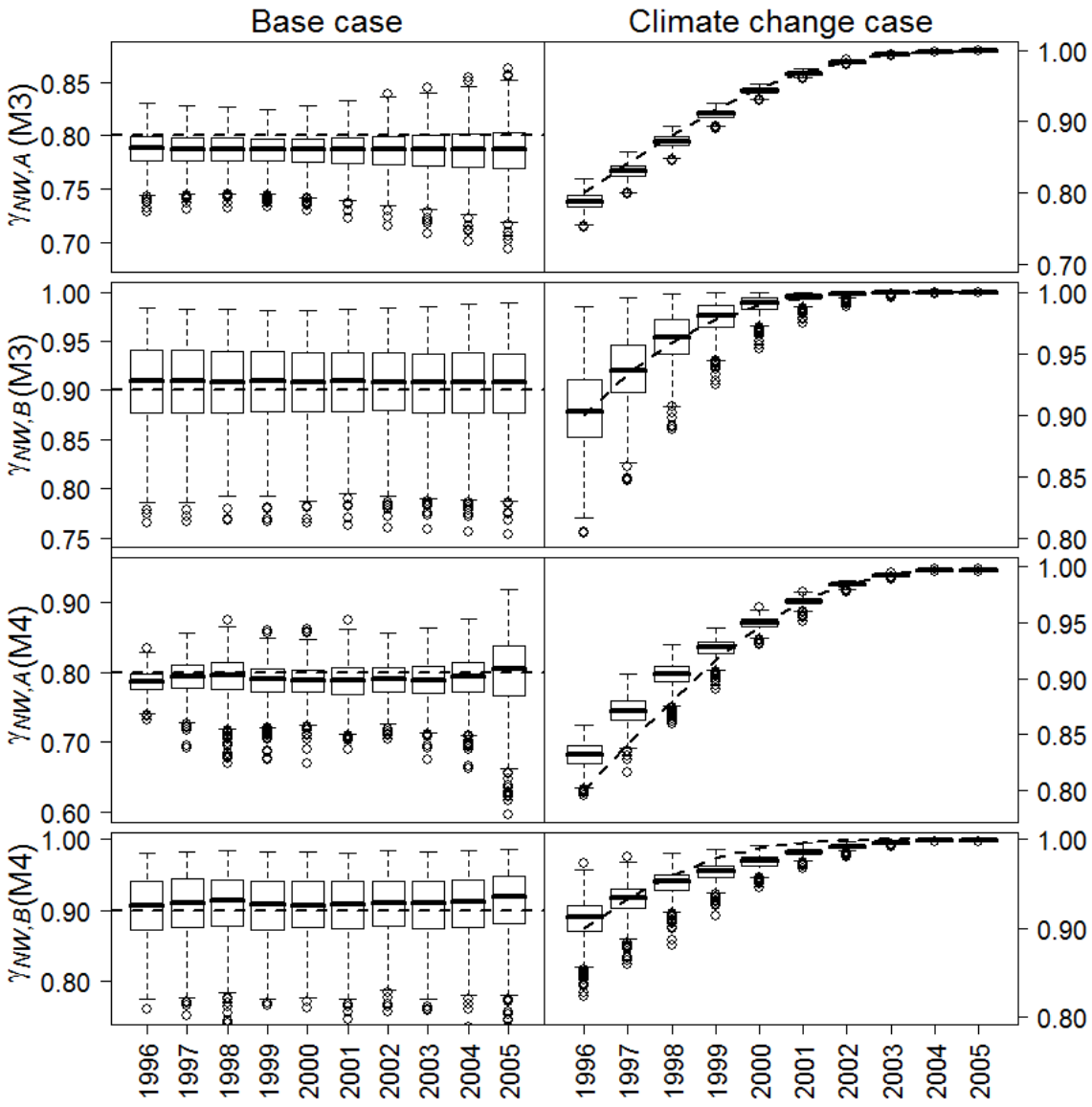

Figure 3.4. Estimates of ontogenetic parameters. Ontogenetic parameters ($\gamma_{R,S}$, *R*-region, *S*-season) estimated from model scenarios M3 and M4 under the base case and climate change case. The dashed lines are the values from the operating model.

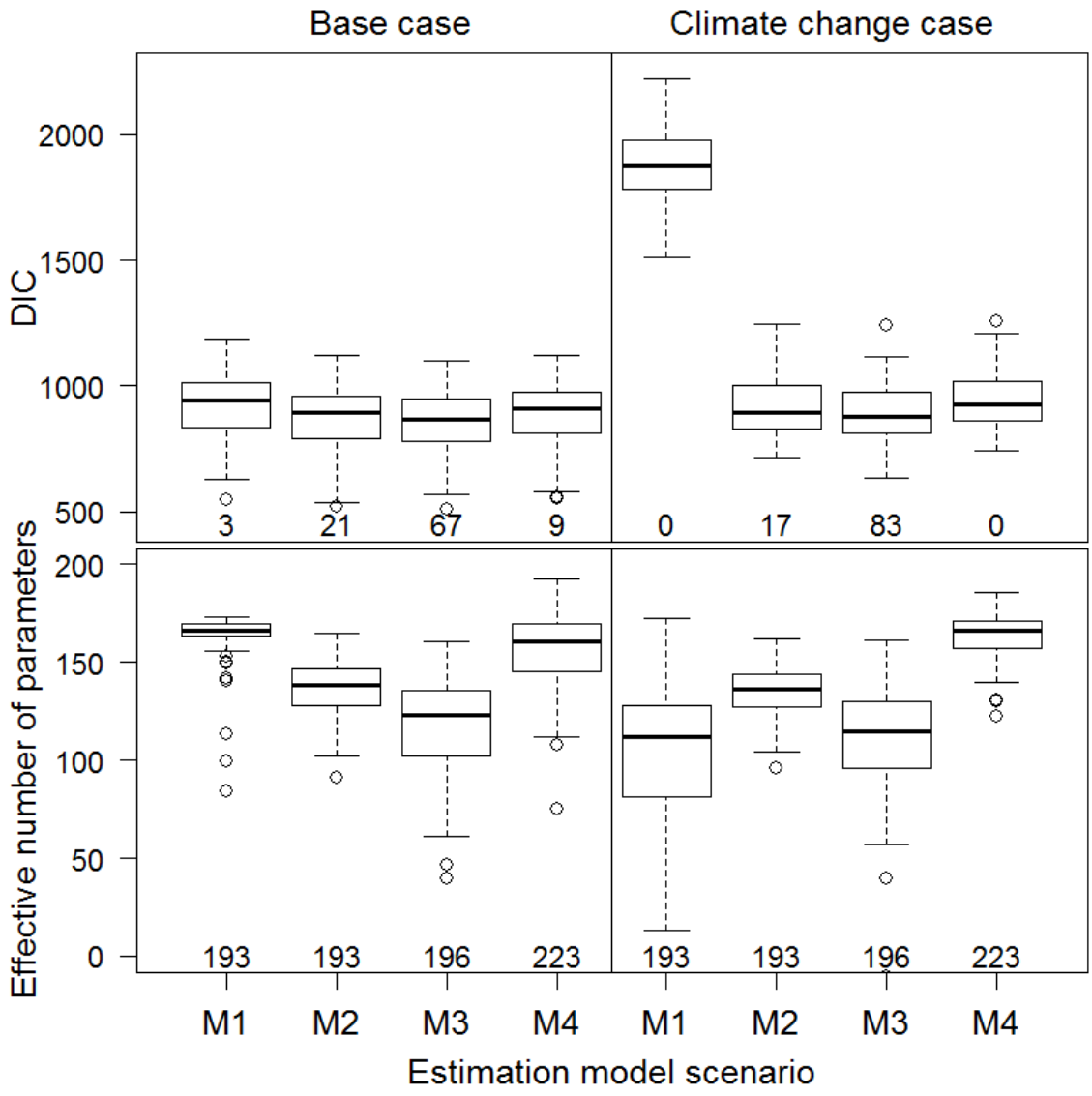

Figure 3.5. Deviance information criterion and effective number of parameters across model scenarios. The numbers shown at the bottom of the top panel are the number of times each particular estimation model scenario had the lowest DIC value. The numbers shown at the bottom of the lower panel are the actual number of parameters estimated in each model scenario.

Appendix 3.A: Markov-Chain Monte Carlo Chain Sampling

3.A.1 Introduction

This appendix details how samples were obtained from the Markov-Chain Monte Carlo (MCMC) simulation to calculate the Deviance Information Criterion (DIC) are provided. To illustrate MCMC chain results across the estimation model scenarios, MCMC chains for model scenario M2 are shown in this appendix. The two primary considerations for obtaining samples from the MCMC chain were chain convergence and autocorrelation.

3.A.2 Methods

For the base case and climate change case, five MCMC chains were simulated with AD Model Builder (ADMB) for model scenario M2. The chain lengths were 10,000, 50,000, 100,000, 500,000, and 1,000,000 saving samples of parameter estimates every 2, 10, 20, 100, and 200 simulations for total sample sizes of 5,000 for each MCMC chain. The standard deviation (SD) in the joint negative log-likelihood value (-lnL) was computed across the MCMC samples of parameter estimates as the MCMC chain grew to illustrate convergence. Autocorrelation among samples taken for each MCMC chain was also computed for the 5 MCMC chains.

3.A.3 Results

For model scenario M2, the MCMC chain converged more quickly and autocorrelation decreased as the length of the chain became larger (Figure 3.A.1 and 3.A.2). At a chain length of 500,000 the SD in –lnL converged by the 2,000th sample for both the base case and climate change case (2nd to bottom panel, Figure 3.A.1). At the same chain length, autocorrelation was reduced to less than 0.1 by a lag of 10 (Figure 3.A.2).

3.A.4 Conclusions

For all model scenarios MCMC chain lengths of 500,000, saving every 100th sample, were simulated to calculate DIC. To ensure the MCMC chain converged the last half of the 5,000 MCMC samples were taken for further analysis. To reduce autocorrelation, a lag of 10 was used to further thin the MCMC samples to 250 samples.

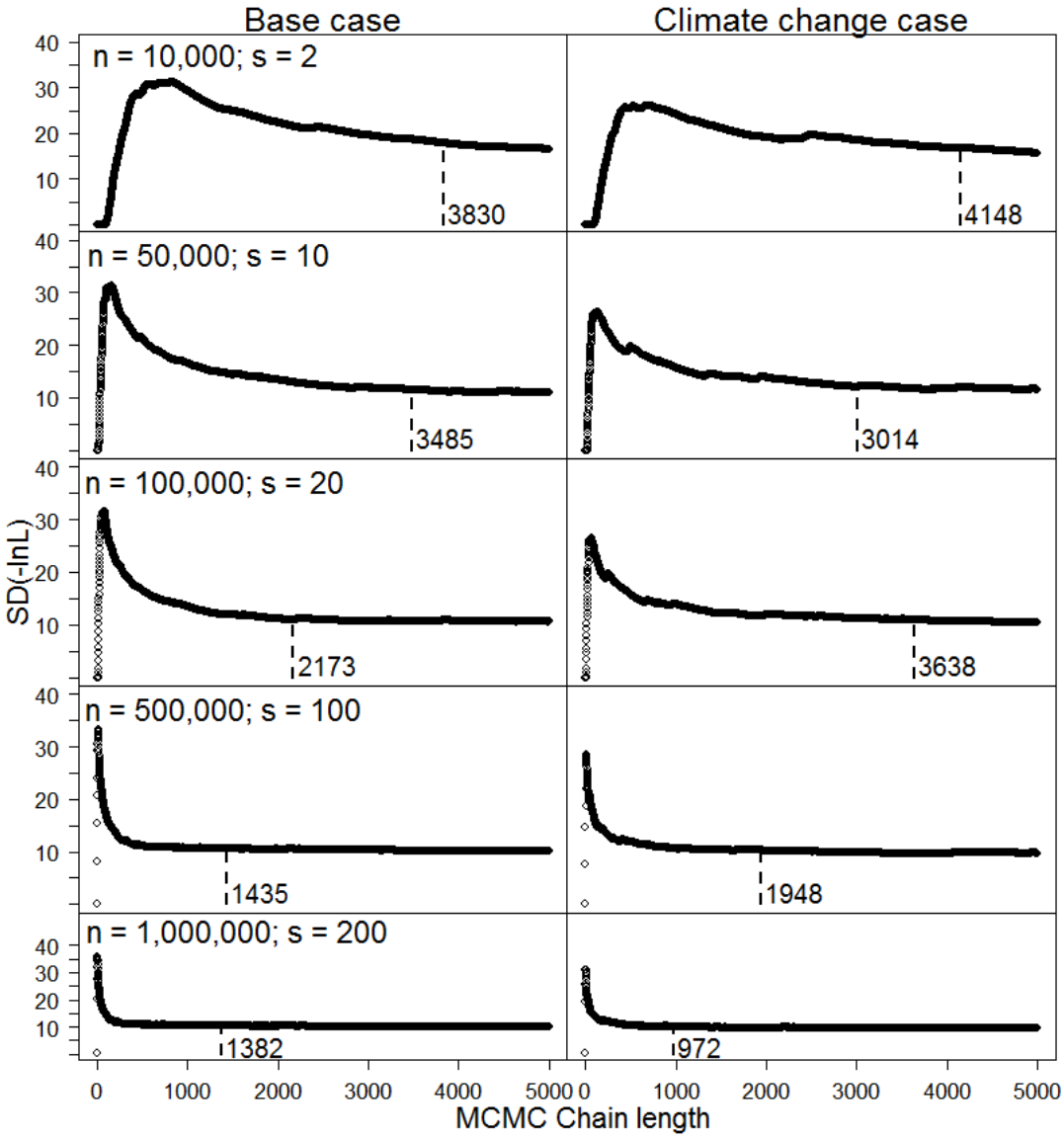

Figure 3.A.1. Standard Deviation in the negative joint log-likelihood. The SD(-lnL) was calculated from the 5 MCMC chain lengths (n-chain length, s-simulation sampled) under the base case and climate change case for model scenario M2. Vertical dashed lines and corresponding number represent the chain length at which the SD in -lnL changed less than 5%.

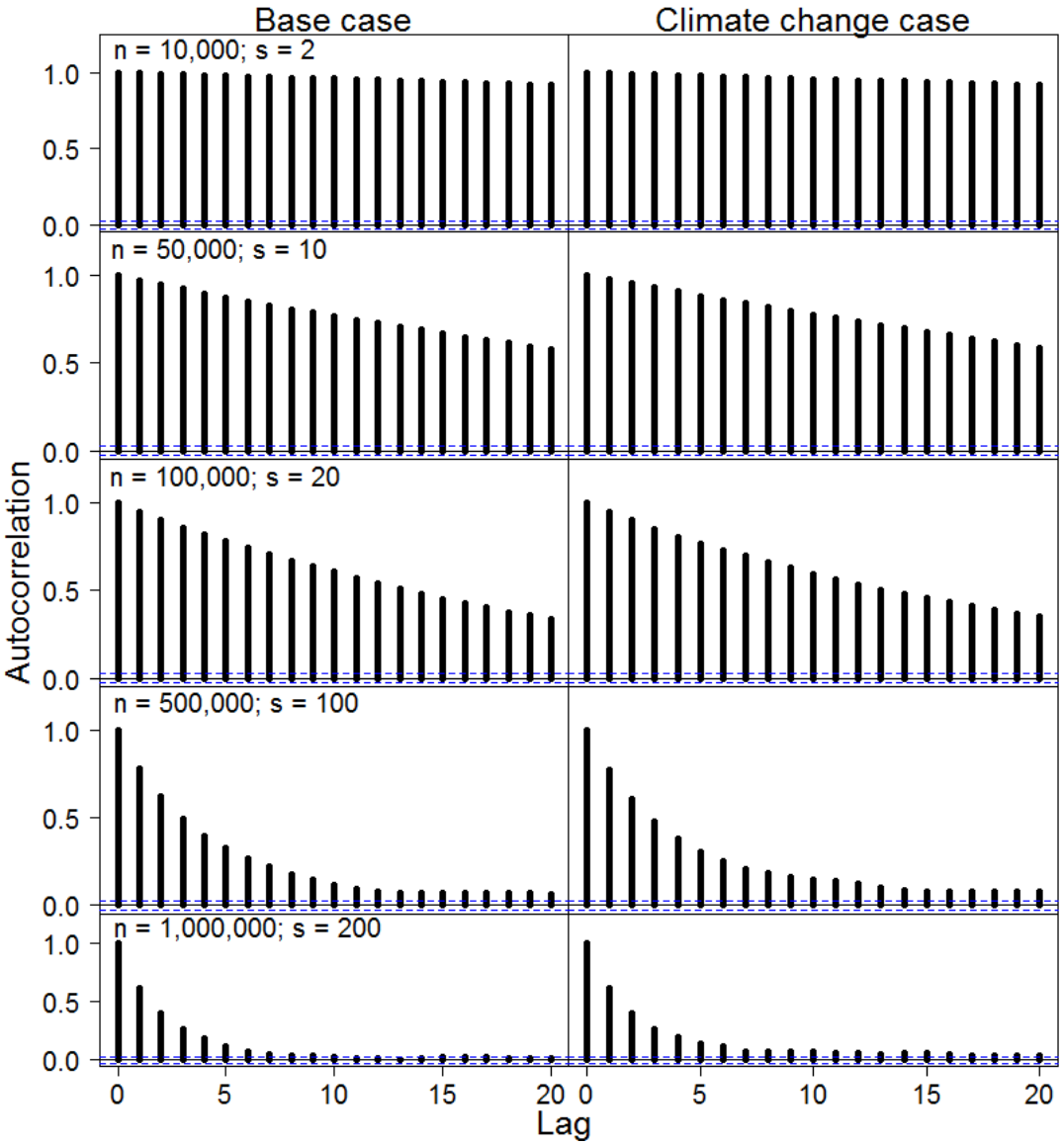

Figure 3.A.2. Autocorrelation of the Standard Deviation in the negative joint log-likelihood. Shown at lags from 0 to 20 among samples from the 5 MCMC chain lengths (n-chain length, s-simulation sampled) under the base case and climate change case for model scenario M2. Horizontal dashed lines represent autocorrelation values of 0.05 and -0.05.

Chapter 4: A Metapopulation Age-Structured Assessment Model for Pacific herring (*Clupea pallasii*) in Southeast Alaska[1]

Abstract

Metapopulation analysis has received increasing attention in the evaluation of fish population dynamics, but it has rarely been applied for the purposes of fisheries management. In this study we develop a metapopulation age-structured assessment model for four Pacific herring (*Clupea pallasii*) populations that spawn in southeast Alaska: Sitka Sound, Craig, Seymour Canal, and Tenakee Inlet. Information about the four Pacific herring populations was integrated into a metapopulation age-structured assessment model through sharing time-dependent deviations in parameters for natural mortality and maturity. The preferred model using the deviance information criterion shared time-dependent deviations in natural mortality between outside populations (Sitka and Craig) and inside populations (Seymour and Tenakee) and shared deviations in maturity among all populations. Further, the metapopulation age-structured assessment model resulted in smaller uncertainty in parameters and population quantities compared to the individual population age-structured assessment models. Due to reduced uncertainty we recommend that the metapopulation age-structured assessment model be considered for future stock assessment of Pacific herring in southeast Alaska and any set of populations for which population parameters can be shared.

[1] Hulson, P.-J. F., Quinn, T. J., II, Hebert, K., and Dressel, S. C. in preparation. A metapopulation age-structured assessment model for Pacific herring (*Clupea pallasi*) in southeast Alaska. In preparation to be submitted to Canadian Journal of Fisheries and Aquatic Sciences.

4.1 Introduction

In fisheries science a large amount of effort has been devoted to identifying key traits that differentiate populations; however, less attention has been focused on discovering traits that link populations across broad spatial scales (Hay *et al.* 2008). It should be noted that differentiation of fish populations has been essential for fisheries management, as fishery effort is mostly regulated for specific areas and fish populations. Progressively more research is being devoted to identify influences on fish populations at broader scales (Williams and Quinn 2000b), although, these studies have focused on evaluating correlation among environmental indices and population dynamics and have rarely been extended to fishery stock assessment for the purposes of management.

Metapopulation analysis has received increasing attention in fisheries science (e.g., Myers *et al.* 1999, Kritzer and Sale 2004). A metapopulation is defined as a group of populations of the same species occupying distinct areas (Levins 1969). The metapopulation model has been extensively utilized in the fields of ecology and evolutionary biology (Smedbol *et al.* 2002). Models that incorporate movement, termed spatially explicit models, have also come to the forefront in fisheries stock assessment (Cadrin and Secor 2009, Goethel *et al.* 2011). However, the metapopulation concept has rarely been applied to model the population dynamics of spatially distinct populations of the same species within stock assessment models used for fisheries management. Rather, stock assessment models that are used for fisheries management either model the overall population as a single group (e.g., Hanselman *et al.* 2010) or model the individual groups within a population separately (e.g., Carlile *et al*. 1999).

Pacific herring (*Clupea pallasii*) in the northern Pacific Ocean provides a unique example of a species that has been investigated across large spatial scales and that may be considered as a metapopulation. For example, Pacific herring populations share similar relationships with environmental conditions across the North Pacific (Haegele and Schweigert 1985, Wespestad 1991, Zebdi and Collie 1995). In addition, latitudinal similarities and differences have been discovered in growth, population diversity, and recruitment for Pacific herring ranging from Oregon to the Bering Sea (Hay *et al.* 2008). Also, Pacific herring populations across the Gulf of Alaska – particularly in Southeast Alaska – can be clustered into statistically significant groups based on similarities in weight at age and spawner-recruit relationships (Williams and Quinn 2000a).

We developed and evaluated a metapopulation age-structured assessment (MPASA) model for several Pacific herring populations in Southeast Alaska. Four populations were considered to have sufficient information for the MPASA model: Pacific herring that spawn in Sitka Sound, Craig, Seymour Canal, and Tenakee Inlet. In an effort to achieve parsimony in the MPASA model we investigated the feasibility of sharing time-dependent deviations in parameters for natural mortality and maturity at age across geographically distinct Pacific herring populations. The goal of this study is to determine whether it is possible to perform stock assessment for several distinct populations within a broad scale assessment model.

4.2 Data

In this study, datasets from four Pacific herring populations in southeast Alaska were included in the metapopulation age-structured assessment model (MPASA, Figure 4.1). The MPASA model integrated four datasets: most with data from 1980-2009 with the exception being Tenakee, for which data were only available from 1982-2009 (see Table 4.1 for time series and notation). The MPASA model was fitted to three main datasets for each Pacific herring population: estimates of total egg deposition from annual spawn surveys, age composition of catch for the commercial fisheries, and age composition from fishery-independent samples of spawning herring. Other population-specific datasets used in the MPASA model included average weight at age of fish captured in the commercial fisheries, time-dependent fecundity at age, and total catch biomass from all participating fisheries. Recently, reader drift was discovered in the ageing of southeast Alaska herring, which affects the catch and spawning age composition datasets, as well as weight at age. The updated data were not available for use in this analysis; we hypothesize that, although parameter estimates would change, the results shown herein would be affected similarly for each population and model structure; thus, the overarching conclusions would be similar.

4.3 Model

The MPASA model was a standard implementation of an age-structured assessment model (ASA, Quinn and Deriso 1999, chapter 8). The original formulation of the individual population ASA models for Pacific herring in southeast Alaska was

developed by Funk and Sandone (1990). The same population dynamics equations from the individual population ASA models were used in the MPASA model; the MPASA model integrated the four populations through sharing time-dependent deviations in parameters and an objective function that included fits to each population's datasets. Population specific parameters estimated within the MPASA model include recruitment of age-3 Pacific herring (in millions of fish, from 1980-2009), initial abundance (in millions of fish, age-4 to 8+ in 1980), natural mortality, logistic parameters for age-specific maturity, logistic parameters for gear selectivity, Ricker spawner-recruit parameters, and mean effective sample size for catch and spawning age compositions. Parameter notation and description is given in Table 4.2.

Pre-fishery abundance ${}^{P}N_{a,t}$, the population size in a year just before spawning and the ensuing commercial fisheries, was estimated by a straightforward survival equation:

$$ {}^{P}N_{a+1,t+1} = \left({}^{P}N_{a,t} - {}^{P}C_{a,t}\right){}^{P}S_{t} \tag{4.1} $$

in which a cohort at age- a+1 and year- t+1 was dependent on the previous year's cohort abundance. The superscript shown prior to the variable represents the population, P. A cohort within the pre-fishery abundance matrix decreased from one year to the next due to catch (${}^{P}C_{a,t}$) and natural mortality. Here ${}^{P}S_{t}$ was the annual natural survival fraction with the subscript t to reflect the time-dependent parameterization. Survival was related to instantaneous natural mortality ($^{P}M_{t}$) by the standard equation (Quinn and Deriso 1999):

$$ {}^{P}S_t = e^{-{}^{P}M_t} \tag{4.2} $$

Catch at age in numbers in the MPASA model was:

$$ {}^{P}C_{a,t} = \frac{{}^{P}\hat{\theta}_{C,a,t}\,{}^{P}Y_t}{\sum_{A} {}^{P}\hat{\theta}_{C,a,t}\,{}^{P}w_{a,t}} \tag{4.3} $$

where ${}^{P}\hat{\theta}_{C,a,t}$ was the estimated catch age composition, ${}^{P}Y_t$ was the total catch biomass, and ${}^{P}w_{a,t}$ was the average weight at age. Estimated catch age composition was:

$$ {}^{P}\hat{\theta}_{C,a,t} = \frac{{}^{P}V_a\,{}^{P}N_{a,t}}{\sum_{a} {}^{P}V_a\,{}^{P}N_{a,t}} \tag{4.4} $$

in which ${}^{P}V_a$ was the scaled gear selectivity, calculated by dividing the logistic function for gear selectivity at age by the maximum gear selectivity so that selectivity at the oldest ages was equal to 1.

Spawning population abundance and biomass were also obtained from the pre-fishery abundance using maturity and subtracting the catch. Spawning population abundance was calculated as:

$$ {}^{P}SN_{a,t} = {}^{P}N_{a,t}\,{}^{P}mat_{a,t} - {}^{P}C_{a,t} \tag{4.5} $$

in which ${}^{P}mat_{a,t}$ was the maturity at age estimated with the logistic function. Spawning biomass was simply the sum of the product of weight at age (${}^{P}w_{a,t}$) and spawning population abundance:

$$ {}^{P}SB_{t} = \sum_{a} {}^{P}w_{a,t}\,{}^{P}SN_{a,t} \tag{4.6} $$

Estimated spawning age composition was also calculated from the spawning population abundance as:

$$ {}^{P}\hat{\theta}_{a,t} = \frac{{}^{P}SN_{a,t}}{\sum_{a} {}^{P}SN_{a,t}} \tag{4.7} $$

The estimated egg deposition was formulated as the product of the female proportion of the population (0.5 assumed from field observations, see also Hulson *et al.* 2008 for proportions of Prince William Sound Pacific herring for coparison), the spawning population abundance, and fecundity:

$$ {}^{P}\hat{E}_t = \sum_a 0.5\,{}^{P}SN_{a,t}\,{}^{P}f_{a,t}\left({}^{P}w_{a,t}\right) \tag{4.8} $$

where ${}^{P}f_{a,t}\left({}^{P}w_{a,t}\right)$ was the fecundity, calculated as an increasing linear function of the average weight at age (${}^{P}w_{a,t}$). For Sitka and Tenakee the fecundity relationship was time-dependent, where the slope and intercept changed in 1984, 1998, and 2002. For Craig and Seymour fecundity was constant from 1980-2009.

The four Pacific herring populations were integrated within the MPASA model through an objective function that included fits to each population's datasets. The MPASA model was fit to egg deposition with the log-normal distribution, and the negative log-likelihood was given by:

$$ -\ln L_{{}^{P}E} = \sum_t \left\{ \ln\left({}^{P}\hat{E}_t\right) + \frac{1}{2}\ln\left(2\pi\,{}^{P}\hat{\sigma}_t^2\right) + \frac{1}{2\,{}^{P}\hat{\sigma}_t^2}\left[\ln\left({}^{P}\hat{E}_t\right) - \ln\left({}^{P}E_t\right)\right]^2 \right\} \tag{4.9} $$

in which annual estimates of variance for each stock (${}^{P}\hat{\sigma}_t^2$) were obtained through resampling methodology presented in Appendix 4.A. The MPASA model was fit to catch and spawning age compositions with the Dirichlet distribution in which mean effective sample size was parameterized (Chapter 2). The negative log-likelihood for catch age composition was:

$$-\ln L_{P_{\theta_C}} = \sum_t \left\{ \sum_a \left[\ln \Gamma\left({}^P\hat{n}_C \, {}^P\hat{\theta}_{C,a,t} \right) - \left({}^P\hat{n}_C \, {}^P\hat{\theta}_{C,a,t} - 1 \right) \ln\left({}^P\theta_{C,a,t} \right) \right] - \ln \Gamma {}^P\hat{n}_C \right\} \quad (4.10)$$

where ${}^P\hat{n}_C$ was the estimated mean effective sample size for the catch age composition, by population-*P*. Similarly, the negative log-likelihood for spawning age composition was:

$$-\ln L_{P_{\theta_{Sp}}} = \sum_t \left\{ \sum_a \left[\ln \Gamma\left({}^P\hat{n}_{Sp} \, {}^P\hat{\theta}_{Sp,a,t} \right) - \left({}^P\hat{n}_{Sp} \, {}^P\hat{\theta}_{Sp,a,t} - 1 \right) \ln\left({}^P\theta_{Sp,a,t} \right) \right] - \ln \Gamma {}^P\hat{n}_{Sp} \right\}$$

(4.11)

where ${}^P\hat{n}_{Sp}$ was the estimated mean effective sample size for the spawning age composition, by population-*P*. The objective function was formulated as the sum of the negative-log-likelihoods for each population's datasets, given by:

$$O = -\sum_P \left(\ln L_{P_E} + \ln L_{P_{\theta_C}} + \ln L_{P_{\theta_{Sp}}} \right) \quad (4.12)$$

4.4 Model scenarios

The model scenarios considered in this study were constructed to evaluate whether a MPASA model that shared time-dependent deviations in parameters across populations was preferred over a model that estimated independent parameters. Following previous investigation into time-dependent parameterization of natural

mortality, maturity, and gear selectivity in the Sitka ASA model (Hulson 2007), both natural mortality and maturity were time-dependent in the MPASA model. Time-dependent gear selectivity was not considered, because including it resulted in no improvement in model performance for the Sitka ASA model (Hulson 2007). Estimating natural mortality and maturity as constant within time blocks was preferred over estimating annual deviations for the Sitka ASA model and were applied to each population in the MPASA model; for natural mortality the time blocks were from 1980-1998 and 1999-2009, and for maturity the time blocks were 1980-2002 and 2003-2009 (Hulson 2007). Time-dependent deviations in parameters were shared across populations by estimating initial, population-specific, parameters for natural mortality and maturity for the first time block, and a shared deviation for the second time block. For natural mortality this was formulated as:

$$^{P}M_{b} = {}^{P}M_{b-1} + \delta_{M} \tag{4.13}$$

in which δ_{M} was the shared deviation for time block-*b*. The estimated parameters in the logistic function for maturity were formulated as:

$$^{P}\alpha_{b}^{m} = {}^{P}\alpha_{b-1}^{m} + \delta_{\alpha^{m}} \tag{4.14}$$

for the age at 50% maturity (inflection parameter), and:

$$ {}^{P}\beta_b^m = {}^{P}\beta_{b-1}^m + \delta_{\beta^m} \tag{4.15} $$

for the slope parameter.

Overall, nine model scenarios were considered in this study for the MPASA model (Table 4.3). Model scenarios M1-M9 were constructed as combinations of sharing time-dependent deviations in parameters for natural mortality and maturity (Table 4.3). Three cases were considered for sharing deviation parameters; either the deviation for both natural mortality and maturity were shared, only one deviation for natural mortality or maturity was shared, or no deviations were shared. We also considered sharing time-dependent deviations in parameters among all populations, or only between inside populations (Seymour Canal and Tenakee Inlet) and outside populations (Sitka Sound and Craig). Inside populations were defined as those within the inner passages of southeast Alaska, and outside populations were defined as those in regions that are open to the Gulf of Alaska (Figure 4.1). Distinction was made between inside and outside populations due to potentially different oceanographic domains and from results of previous analysis that suggested more similarity between populations within the outside and inside domains (Williams and Quinn 2000a). Model scenarios M1-M9 are presented in order of increasing number of parameters. For model scenario M9, none of the parameters were shared among populations so this model scenario was just a combination of the individual ASA models within the MPASA model. The main difference between model scenario M9 and the individual ASA models was that the time blocks for natural

mortality and maturity were considered to be the same among populations, allowing for more direct comparison with the other MPASA model scenarios (for the individual population time blocks see Appendix 4.B). Model scenario M9 was considered so that model comparison criteria could be used to compare between the individual population ASA models and the MPASA model.

Model comparison among MPASA model scenarios M1-M9 was performed with the Deviance Information Criterion (DIC, Spiegelhalter *et al.* 2002). For each model scenario the Markov-Chain Monte Carlo (MCMC) simulation chain length was 10,000,000 and, to ensure chain convergence and reduce autocorrelation, the last half of samples obtained from sampling every 5,000 steps in the MCMC chain were selected for DIC calculation, resulting in 1,000 samples. DIC was computed as (following Spiegelhalter *et al.* 2002):

$$DIC = \overline{D(\Theta)} + p_D \tag{4.16}$$

where $D(\cdot)$, the 'Bayesian deviance' was defined as:

$$D(\Theta) = -2\ln L(y|\Theta) + 2\ln L(y) \tag{4.17}$$

in which $L(y|\Theta)$ is the likelihood function for the data, *y*, given the estimated parameters, Θ and $\ln L(y)$ was the likelihood value when the predicted values equaled the observed data. The effective number of parameters, p_D, was calculated as:

$$p_D = \overline{D(\Theta)} - D(\overline{\Theta}) \tag{4.18}$$

where $\overline{D(\Theta)}$ was the average deviance across the MCMC samples of parameter estimates, and $D(\overline{\Theta})$ was the deviance at the average of the parameter estimates across the MCMC samples.

4.5 Results

The smallest DIC value resulted from MPASA model scenario M2, in which time-dependent deviation parameters for natural mortality were shared within the inside (Seymour and Tenakee) and outside (Sitka and Craig) populations and time-dependent deviation parameters for maturity were shared among all the populations (Table 4.4). The next smallest DIC value resulted from model scenario M6, in which no deviation parameters were shared for natural morality among populations and time-dependent deviations were shared for maturity within the inside and outside populations (Table 4.4). All the model scenarios that shared time-dependent deviation parameters for natural mortality and/or maturity in some combination (M1-M8) resulted in smaller DIC values than model scenario M9, in which none of the time-dependent deviation parameters were

shared among populations (Table 4.4). The smallest effective number of parameters resulted from model scenario M2, the second smallest from model scenario M3, and the largest effective number of parameters resulted from model scenario M9 (Table 4.4). Because model scenario M2 resulted in the smallest DIC value, this model scenario was selected for further comparison with the results of individual ASA models from model scenario M9.

In both of the individual ASA models (model scenario M9) and the preferred MPASA model scenario (M2), parameter estimates of natural mortality decreased after 1999 and parameters for the age at 50% maturity increased after 2003 (Table 4.5). The only exception to the change in age at 50% maturity was seen in the Seymour population in model scenario M9 (Table 4.5). The coefficient of variation (CV) in natural mortality was smaller both pre – 1998 and post – 1999 in model scenario M2 compared to M9 for Sitka and Craig, and was larger in M2 compared to M9 for Seymour and Tenakee post – 1999 (Table 4.5). There was a smaller CV for parameter estimates of the age at 50% maturity from model scenario M2 compared to M9 for all populations and both time periods with the exception of Seymour and Tenakee pre – 2002.

Estimated spawning biomass for the four Pacific herring populations integrated in the MPASA model was similar in trend and magnitude between model scenarios M2 and M9 (Figure 4.2). For Sitka, estimated spawning biomass from model scenario M9 was within the 95% confidence intervals for model scenario M2 (Figure 4.2). The same result was obtained for estimated spawning biomass from Craig and Tenakee. The estimated spawning biomass for Seymour in 2003 from model scenario M9 was outside of the 95%

confidence intervals from model scenario M2 (Figure 4.2). For the majority of years the CV in estimated spawning biomass was smaller in model scenario M2 compared to M9 (Figure 4.3): 24 of 30 years for Sitka, 21 of 30 years for Craig, 27 of 30 years for Seymour, and 23 of 28 years for Tenakee (Figure 4.3). Further, the decreases in CV for estimated spawning biomass from model scenario M2 compared to M9 were small (maximum decrease around 6%), but for the majority, model scenario M2 had greater precision than model scenario M9.

4.6 Discussion

For the southeast Alaska Pacific herring populations considered in this study, the MPASA model that shared time-dependent deviations in parameters for natural mortality and maturity provided an improvement over the individual ASA models used for stock assessment. Primarily, this improvement was a reduction in model uncertainty; the MPASA model indicated smaller CVs for the majority of parameter and population quantities compared to the individual ASA models. Further, the estimated parameters and population quantities were similar between the MPASA and individual ASA models, with few significant differences in spawning biomass. Due to the reduction in uncertainty with the MPASA model and the similarities between estimates obtained from the MPASA and individual ASA models, we recommend that the MPASA model be considered for future management of Pacific herring populations in Southeast Alaska.

The effective number of parameters resulting from the preferred MPASA model (model scenario M2) was substantially less than that of the individual ASA models

(model scenario M9). Also, the DIC value for model scenario M2 was smaller, adding support to the conclusion that the MPASA model provides improvement over the individual stock ASA models. For all the MPASA model scenarios considered, the effective number of parameters was smaller than the actual number of parameters; the largest reduction in the effective number of parameters compared to the actual number of parameters resulted from the preferred MPASA model scenario.

When sharing parameter deviations in natural mortality, the preferred MPASA model estimated shared deviations for Pacific herring populations that spawn in the inside passage of Southeast Alaska and shared deviations for stocks that spawn in the outside regions that are open to the Gulf of Alaska. Previous analysis suggests that correlation groupings based on recruitment and weight at age among Pacific herring populations in the North Pacific Ocean follow the same pattern; that stocks located within inside passages are more closely related to one another than to stocks whose spawning grounds are open to the Gulf of Alaska (Zheng 1996, Williams and Quinn 2000a). Mechanisms for such groupings are difficult to discern (e.g., Williams and Quinn 2000b) and may be related to oceanographic conditions that differ between the inside and outside domains (e.g., Zebdi and Collie 1995), possibly affecting food availability or predator abundance. We also found that sharing parameter deviations for maturity across all the populations was preferred over parameterizing based on inside or outside locations. This result may indicate broad-scale influences on maturity (e.g., Hay *et al.* 2008) coupled with localized influences on natural mortality.

With the development of a MPASA model for Pacific herring in southeast Alaska, further investigations into broad and localized scale influences on population dynamics are possible. For example, rather than investigating environmental relationships within ASA models (e.g., Hulson 2007) or spawner-recruit models (e.g., Zebdi and Collie 1995) for a single population, it would be possible to include environmental covariates into the MPASA model that influence several populations simultaneously. These same methods would also be easily extended to the evaluation of influences on mortality such as from disease (e.g., Marty *et al.* 2010) or predator abundance (e.g., Deriso *et al.* 2007). Further, in this study we only included four Pacific herring populations into the MPASA model. Following the methods detailed here a number of Pacific herring populations both in southeast Alaska (e.g., Hoonah Sound, Ernest Sound, etc.) and out to the Gulf of Alaska (e.g., Kodiak, Prince William Sound) could be included and broad scale investigations would be possible.

The four Pacific herring populations integrated in the MPASA model are geographically close, thus exchange among the populations through movement is possible (e.g., Hay *et al.* 2001). The main caveat in the MPASA model is that we did not consider estimation of movement, primarily because no data are available to estimate movement. If movement and exchange between these populations occur, the movement within the MPASA model would be confounded with parameters for mortality and recruitment, in which case tagging studies might help address the parameter confounding (e.g., Anganuzzi 1996). Although, this caveat is not unique to the MPASA model, movement is also not considered in the individual population ASA models that are

currently used to manage these herring populations. Further, all assumptions on parameters and population dynamics are the same between the individual ASA models and the MPASA model. The only additional attribute of the MPASA model compared to the individual ASA models is the sharing of time-dependent parameter deviations, i.e., changes in parameters over time are the same among or between populations.

The MPASA model in this study focused on Pacific herring in southeast Alaska; however, the methods presented here could be extended to stock assessment of other species, or among closely related species. Spatial allocation of catch quotas has long had important biological and socioeconomic implications for fisheries management (e.g., Heifetz *et al.* 1997, Hare 2010) and a MPASA model could provide a way to provide alternative spatial management advice. It would then be possible to provide management advice based on spatial considerations through sharing parameters or time-dependent parameter deviations within a MPASA model.

4.7 Acknowledgments

This publication is the result of research sponsored by Alaska Sea Grant with funds from the National Oceanic and Atmospheric Administration (NOAA) Office of Sea Grant, Department of Commerce, under grant NA06OAR4170013 (project number R/31–16), and from the University of Alaska with funds appropriated by the state. We thank Milo Adkison, Brenda Norcross, Dana Hanselman, Gary Marty, and Dave Carlile for helpful comments and advice, and Kalei Shotwell for the map illustration of the Pacific herring populations considered in this study. The findings and conclusions in this paper

are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA, or the Alaska Department of Fish and Game.

4.8 References

Anganuzzi, A. 1996. Some technical issues regarding the study of fisheries interactions. *In* Status of Interactions of Pacific Tuna Fisheries in 1995, pp. 21-37. Ed. by R. Shomura, J. Majkowski, and R. Harman. FAO Fisheries Technical Paper no. 365.

Cadrin, S. X., and Secor, D. H. 2009. Accounting for spatial population structure in stock assessment: past, present, and future. Fish and Fisheries Series, 31: 405–426.

Carlile, D. W., Larson, R. C., and Hebert, K. P. 1999. Stock assessments of Southeast Alaska herring in 1997 and forecasts for 1998 abundance. Alaska Department of Fish and game, Division of Commercial Fisheries, Regional Information Report 1J99-10, Juneau, Alaska.

Deriso, R. B., Maunder, M. N., and Skalski, J. R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Canadian Journal of Fisheries and Aquatic Sciences, 64: 187–197.

Efron, B., and Tibshirani, R. J. 1993. An introduction to the Bootstrap. Chapman and Hall, New York.

Funk, F. C. and Sandone, G. J. 1990. Catch-age analysis of Prince William Sound, Alaska, herring, 1973-1988. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J89-02, Juneau, Alaska.

Goethel, D. R., Quinn, T. J., II, and Cadrin, S. X. 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Reviews in Fisheries Science, 19(2): 119–136.

Haegele, C. W. and Schweigert, J. F. 1985. Distribution and characteristics of herring spawning grounds and description of spawning behavior. Canadian Journal of Fisheries and Aquatic Sciences, 42: 39–55.

Hanselman, D. H., Lunsford, C. R., and Rodgveller, C. J. 2010. Assessment of the sablefish stock in Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska, pp. 329-468. North Pacific Fishery Management Council, Anchorage, AK.

Hare, S. 2010. Assessment of the Pacific halibut stock at the end of 2010. International Pacific Halibut Commission Report of Assessment and Research Activities 2010.

Hay, D. E., Rose, K. A., Schweigert, J., and Megrey, B. A. 2008. Geographic variation in North Pacific herring populations: Pan-Pacific comparisons and implications for climate change impacts. Progress in Oceanography, 77: 233–240.

Hay, D. E., McCarter, P. B., and Daniel, K. S. 2001. Tagging of Pacific herring *Clupea pallasii* from 1936-1992: a review with comments on homing, geographic fidelity, and straying. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1356–1370.

Hebert, K. 2009. Southeast Alaska 2009 herring stock assessment surveys. Alaska Department of Fish and Game, Division of Sport Fish and Commercial Fisheries, Fishery Data Series No. 09-72, Juneau, Alaska. 102 pp.

Heifetz, J., Fujioka, J. T., Quinn, T. J., II. 1997. Geographic apportionment of sablefish, *Anoplopoma fimbria*, harvest in the northeastern Pacific Ocean. *In* Proceedings of the International Symposium on the Biology and Management of Sablefish. Ed. by M. Saunders and M. Wilkins. pp 229-238. NOAA Tech. Rep. 130.

Hulson, P.-J. F., 2007. Analysis and comparison of age-structured assessment models for two Pacific herring populations. M.S. Thesis. University of Alaska Fairbanks, School of Fisheries and Ocean Sciences. 127 pp.

Hulson, P.-J. F., Miller, S. E., Quinn II, T. J., Marty, G. D., Moffitt, S. D., and Funk, F. 2008. Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. ICES Journal of Marine Science, 65: 25–43.

Kritzer, J. P., and Sale, P. F. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries, 5: 131–140.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America, 15: 237–240.

Marty, G. D., Hulson, P.-J. F., Miller, S. E., Quinn, T. J., II, Moffitt, S. D., and Merizon, R. A. 2010. Failure of population recovery in relation to disease in Pacific herring. Diseases of Aquatic Organisms, 90: 1–14.

Miller, S., Ianelli, J., and Quinn, T. J., II, 2008. Estimation of age-specific migration in an age-structured model. *In* Resiliency of Gadid Stocks to Fishing and Climate Change, pp. 161-178. Ed. by G. Kruse, K. Drinkwater, J. N. Ianelli, J. S. Link, D. L. Stram, V. Wespestad, and D. Woodby. Alaska Sea Grant College Program Report AK-SG-08-01, University of Alaska, Fairbanks.

Myers, R. A., Bowen, K. G., and Barrowman, N. J. 1999. The maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences, 56:1290–1314.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

R Development Core Team 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Schnute, J. T., and Haigh, R. 2007. Compositional analysis of catch curve data, with and application to *Sebastes maliger*. ICES Journal of Marine Science, 64(2): 218–233.

Smedbol, R. K., McPherson, A., Hansen, M. M., and Kenchington, E. 2002. Myths and moderation in marine 'metapopulations'? Fish and Fisheries, 3: 20–35.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, 64(4): 583–639.

Wespestad, V. G., 1991. Pacific herring population dynamics, early life history, and recruitment variation relative to eastern Bering Sea oceanographic factors. Doctoral dissertation, University of Washington, Seattle. 234 pp.

Williams, E. H., and T. J. Quinn II, 1998. A parametric bootstrap of catch-age compositions using the Dirichlet distribution. *In* Fishery stock assessment models. Ed. by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, and C. I. Zhang. Alaska Sea Grant Program Report No. AK-SG-98-01, University of Alaska, Fairbanks.

Williams, E. H., and Quinn, T. J., II, 2000a. Pacific herring, *Clupea pallasii*, recruitment in the Bering Sea and north-east Pacific Ocean, I: relationships among different populations. Fisheries Oceanography, 9(4): 285–299.

Williams, E. H., and Quinn, T. J., II, 2000b. Pacific herring, *Clupea pallasii*, recruitment in the Bering Sea and north-east Pacific Ocean, II: relationships to environmental variables and implications for forecasting. Fisheries Oceanography, 9(4): 300–315.

Zebdi, A. and Collie, J. S. 1995. Effect of climate on herring (*Clupea pallasii*) population dynamics in the Notheast Pacific Ocean, p. 277-290. *In* Climate change and northern fish populations. Ed. by R.J. Beamish Canadian Special Publications of Fisheries and Aquatic Sciences, 121: 277–290.

Zheng, J. 1996. Herring population-recruitment relationships and recruitment patterns in the North Atlantic and Northeast Pacific Oceans. Fisheries Research, 26: 257–277.

Table 4.1. Datasets used in the MPASA model. Datasets used included Sitka Sound (SIT), Craig (CRG), Seymour Canal (SEY), and Tenakee Inlet (TEN) Pacific herring.

Notation	Description
	Datasets used in Objective Function:
${}^{P}\theta_{C,a,t}$	Commercial catch age composition for population-*P*, age-*a*, year-*t* SIT: 1980-2009 CRG: 1980, 1982-1983, 1986-2009 SEY: 1981, 1984, 1986-1990, 1994-1995, 1998-2009 TEN: 1982-1990, 1997-2004, 2009
${}^{P}\theta_{Sp,a,t}$	Spawning age composition for population-*P*, age-*a*, year-*t* SIT: 1980-2009 CRG: 1981, 1984, 1988-2009 SEY: 1980-2009 TEN: 1982-1990, 1993, 1997-2005, 2007-2009
${}^{P}E_{t}$	Egg deposition survey for population-*P* in year-*t* (in trillions of eggs) SIT: 1980-2009 CRG: 1980, 1982-2009 SEY: 1980-2009 TEN: 1984-2009
	Other Datasets used for estimation:
${}^{P}w_{a,t}$	Weight at age for population-*P*, age-*a*, year-*t* (in grams)
${}^{P}f_{a,t}$	Fecundity relationship for population-*P* at age-*a* in year-*t*
${}^{P}C_{a,t}$	Catch in the commercial fisheries at age-*a* in year-*t* (millions of fish)
${}^{P}Y_{t}$	Yield in the commercial fisheries for population-*P* in year-*t* (metric tonnes)

Table 4.2. Parameters estimated in the MPASA model.

Notation	Description
${}^{P}N_{3,t}$	age-3 abundance for population-*P* in year-*t* (recruitment)
${}^{P}N_{a,1980}$	initial abundance for population-*P* in 1980 (Tenakee in 1982)
${}^{P}M_{t}$	natural mortality for population-*P* in year-*t*
${}^{P}\alpha_{t}^{m}$	age at 50% maturity for population-*P* in year-*t*
${}^{P}\beta_{t}^{m}$	slope of maturity-at-age logistic function for population-*P* in year-*t*
${}^{P}\alpha^{g}$	age at 50% fishing selectivity for population-*P*
${}^{P}\beta^{g}$	slope of fishing selectivity-at-age logistic function for population-*P*
${}^{P}\alpha^{SR}$	Ricker spawner-recruit productivity parameter for population-*P*
${}^{P}\beta^{SR}$	Ricker density-dependent parameter for population-*P*
${}^{P}\hat{n}_{C}$	Catch age composition effective sample size for population-*P*
${}^{P}\hat{n}_{Sp}$	Spawning age composition effective sample size for population-*P*

Table 4.3. Description of model scenarios evaluated in the MPASA model. Model scenarios are presented in order of increasing number of parameters. For shared deviations in parameters, 'All' represents model scenarios that shared the deviation parameter among all populations, 'Inside/Outside' represents model scenarios that shared the deviation parameter between populations in the geographic inside (Seymour and Tenakee) or outside domains (Sitka and Craig), and 'None' describes model scenarios that did not share a deviation parameter across populations.

Model Scenario	Stocks for which mortality deviation shared	Stocks for which maturity deviation shared	Total number of parameters
M1	All	All	177
M2	Inside/Outside	All	178
M3	All	Inside/Outside	179
M4	Inside/Outside	Inside/Outside	180
M5	None	All	180
M6	None	Inside/Outside	182
M7	All	None	183
M8	Inside/Outside	None	184
M9	None	None	186

Table 4.4. MPASA model comparison statistics. Deviance information criterion (DIC) values, effective number of parameters, and actual number of parameters resulting from MPASA model scenarios M1 – M9. Model scenarios are presented in order of increasing DIC values.

Model Scenario	Derviance Information Criterion (DIC)	Effective number of parameters (p_D)	Actual number of parameters (p)
M1	433	128	177
M2	412	99	178
M3	425	111	179
M4	424	121	180
M5	426	133	180
M6	419	116	182
M7	430	121	183
M8	436	135	184
M9	438	153	186

Table 4.5. Parameter estimates and coefficient of variation. Shown for natural mortality (pre – 1998 and post – 1999) and age at 50% maturity (pre – 2002 and post – 2003) from the individual population ASA models (M9) and the preferred MPASA model (M2) for each of the four Pacific herring populations considered.

		Parameter value		Coefficient of variation	
	Stock	M2	M9	M2	M9
Natural Mortality pre-1998	Sitka	0.57	0.59	6.5%	6.6%
	Craig	0.74	0.72	8.0%	9.5%
	Seymour	0.54	0.59	9.7%	8.4%
	Tenakee	0.87	0.81	8.0%	9.8%
Natural mortality post-1999	Sitka	0.19	0.18	17%	18%
	Craig	0.36	0.41	15%	15%
	Seymour	0.23	0.22	22%	20%
	Tenakee	0.56	0.54	12%	11%
Age at 50% maturity pre-2002	Sitka	3.7	3.7	5.0%	5.0%
	Craig	4.2	4.4	6.9%	8.1%
	Seymour	4.3	4.6	7.3%	6.0%
	Tenakee	5.0	4.8	6.5%	6.4%
Age at 50% maturity post-2003	Sitka	4.2	4.5	4.9%	5.6%
	Craig	4.7	4.4	7.1%	7.3%
	Seymour	4.8	4.0	7.9%	9.4%
	Tenakee	5.5	6.0	5.6%	7.1%

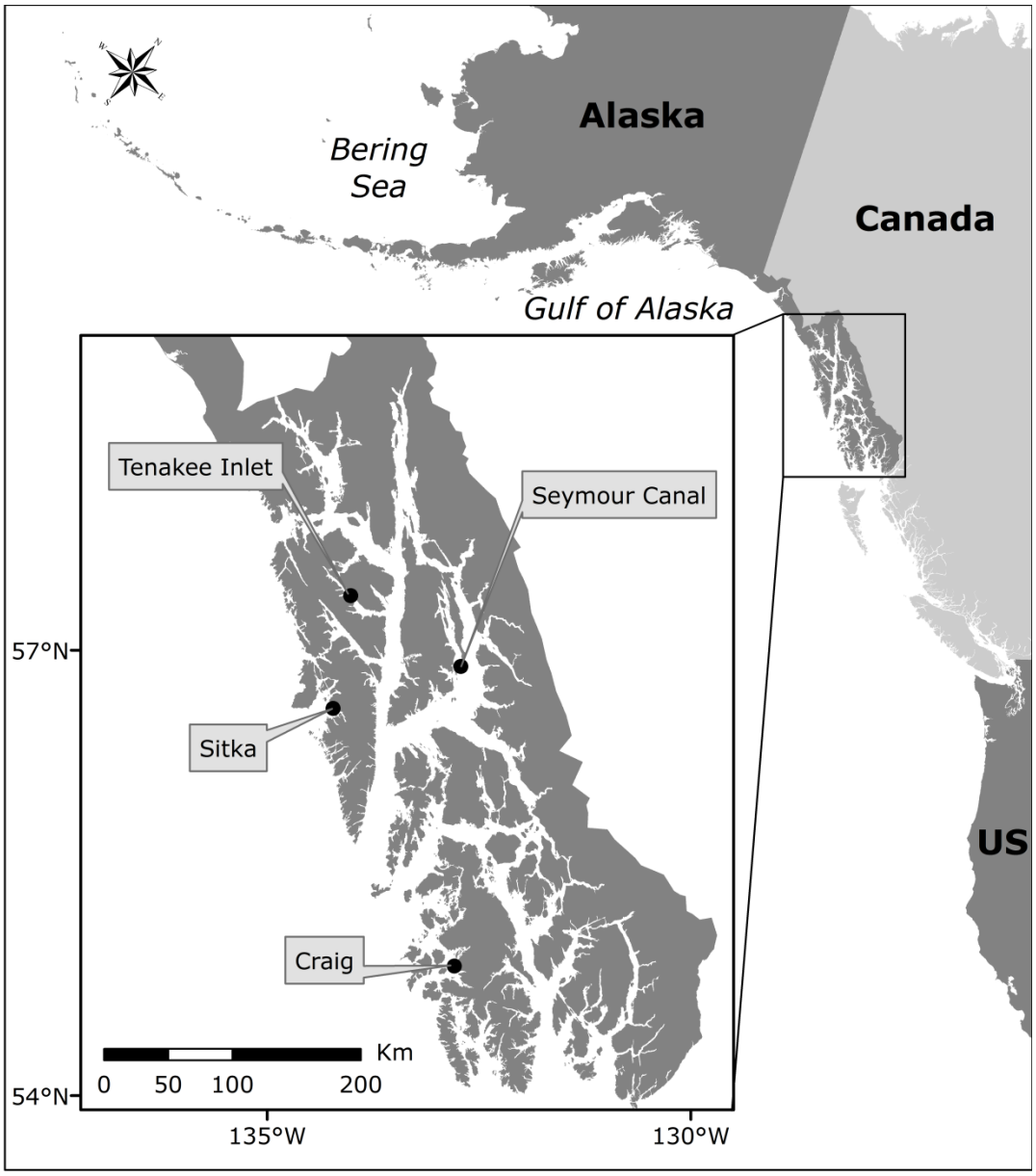

Figure 4.1. Map of the North Pacific Ocean noting the location of the four Pacific herring stocks considered in the MPASA model in SE Alaska.

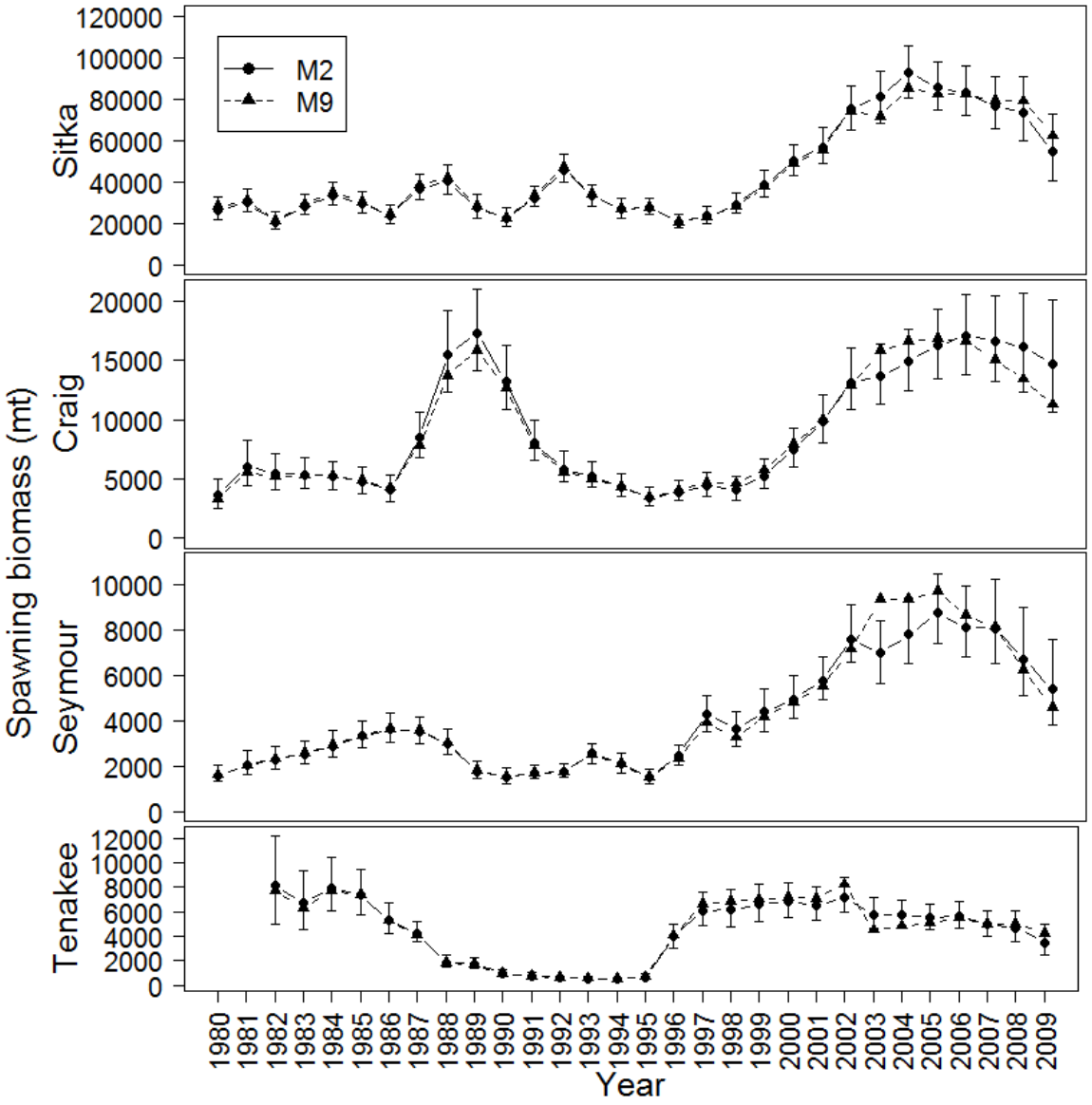

Figure 4.2. Estimated spawning biomass. Shown for the preferred MPASA model scenario (M2; shown with 95% credible intervals) and the individual stock assessment models (M9) for the four Pacific herring populations integrated in the MPASA model.

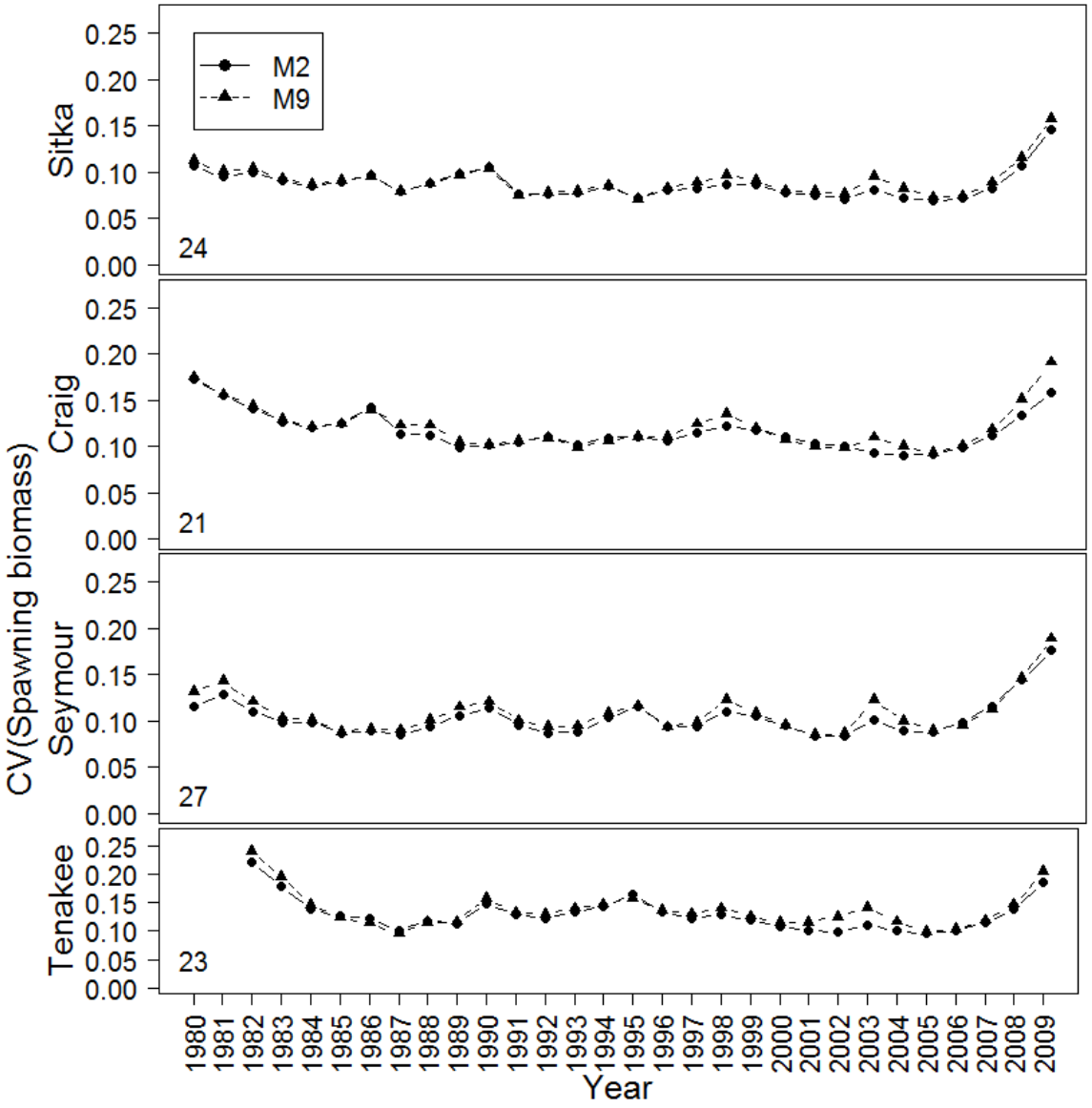

Figure 4.3. Coefficient of variation (CV) in estimated spawning biomass. Shown for the selected MPASA model scenario (M2) and the individual stock assessment models (M9) for the four Pacific herring populations integrated in the MPASA model. The integer shown in the bottom left in each panel is the number of years for which the CV in spawning biomass from M2 was smaller than M9.

Appendix 4.A: Estimation of standard error in annual egg deposition

4.A.1 Introduction

Within this appendix the methodology employed to estimate standard error (SE) in the observations of annual egg deposition is presented. The estimates of SE are subsequently used as estimates of standard deviations within the log-normal likelihood function to fit egg deposition by the individual age-structured assessment (ASA) models and the metapopulation age-structured assessment (MPASA) model for Pacific herring populations that spawn in Sitka Sound, Craig, Seymour Canal, and Tenakee Inlet. Overall, estimation of SE follows from resampling quadrate counts obtained from the two-stage random transect – systematic quadrate sampling design used by the Alaska Department of Fish and Game (ADF&G) in the annual egg deposition survey.

4.A.2 Data – Methods

Raw quadrat count data by transect were collected for the four Pacific herring populations considered in this study from the annual egg deposition surveys conducted by ADF&G. The years of egg deposition fit in the MPASA model for each stock are shown in Table 4.1. The years for which raw data were available included 1991-2009 for Sitka Sound and 1999-2008 for Craig, Seymour Canal, and Tenakee Inlet (Table 4.A.1).

The sampling design used by ADF&G to collect annual observations of egg deposition is a two-stage random transect – systematic quadrat design (Hebert 2009). Through the course of spawning, aerial surveys are conducted to map the beaches upon

which Pacific herring spawn. After spawning has ended, transects (perpendicular to the beach) are randomly selected within the areas that received spawn, and the number of transects performed follow from considerations of precision (Hebert 2009). Counts of egg deposition are obtained from placing 0.1 meter^2 quadrats in 5 meter intervals along transects until no eggs are present; these counts are then corrected for each diver through double sampling (Hebert 2009, Jessen 1978). Total egg deposition in the survey is estimated as:

$$E_t = A\overline{\rho} \tag{4.A.1}$$

in which A is the estimated total area (meter^2) on which eggs have been deposited and $\overline{\rho}$ is the estimated mean density of eggs per quadrat. The total area A is estimated as:

$$A = L\overline{w} \tag{4.A.2}$$

where L is the total length of shoreline receiving spawn and $\overline{w}$ is the mean width of spawn, estimated by the mean length of transects. The mean density of eggs $\overline{\rho}$ is estimated as:

$$\overline{\rho} = 10 \cdot \left[\frac{\sum_q v_q}{m} \right] \tag{4.A.3}$$

in which v_q is the corrected visual estimate of egg numbers in quadrat q and m is the total number of quadrats. Because egg estimates are made within 0.1 meter^2 quadrats, multiplying by 10 expresses $\bar{\rho}$ in eggs per meter^2.

The resampling methodology employed to estimate the SE in the annual egg deposition survey involved resampling of transect and quadrat counts. In the first step of resampling, transects are randomly sampled with replacement. In the second step, all of the quadrat corrected egg counts within a transect are pooled, then randomly sampled with replacement out to the total length of each individual transect. Randomly sampling transects essentially defines the total number of quadrats that are resampled, as the length of transects (i.e., a function of the number of quadrats within each transect) is variable. Egg deposition estimates are then calculated by equations 4.A.1 – 4.A.3 across 1000 replications of the egg deposition survey for each stock to obtain robust estimates of SE (e.g. Efron and Tibshirani 1993). Resampling of transect – quadrat data was performed with code developed in the programming language R (e.g. R Development Core Team 2008).

4.A.3 Results – Conclusions

Overall, estimates of SE were comparable among populations, with the exception of a large SE for Sitka in 2008 (Table 4.A.1). No relationship was found between the log transformed egg deposition estimates and the estimated SE from the resampling methodology (Figure 4.A.1). The mean estimated SE across the years for which egg

deposition variance was estimated was smaller for Sitka than for the other populations and was comparable between Craig, Seymour, and Tenakee, likely due to the larger number of transects performed in Sitka compared to the other populations (horizontal lines in Figure 4.A.1).

Within the MPASA model and individual ASA models, the annual estimates of SE for the egg deposition datasets were used in the log-normal likelihood to weight individual years. Because there was a lack of a strong relationship between the annual log transformed egg deposition and estimated SE, the mean estimated SE was used for years in which raw transect – quadrat data was not available (prior to 1991 for Sitka, and prior to 1999 for Craig, Tenakee, and Seymour).

Table 4.A.1. Annual egg deposition with associated estimates of standard error. Shown for years in which raw data were available for the four Pacific herring populations considered in the MPASA model.

	Sitka Sound		Craig		Seymour Canal		Tenakee Inlet	
Year	E_t	SE(E_t)	E_t	SE(E_t)	E_t	SE(E_t)	E_t	SE(E_t)
1991	2.35	0.24						
1992	4.34	0.22						
1993	3.72	0.22						
1994	1.49	0.19						
1995	3.50	0.22						
1996	3.59	0.20						
1997	2.89	0.26						
1998	3.58	0.19						
1999	4.50	0.24	0.56	0.30	0.33	0.30	0.97	0.29
2000	5.85	0.24	0.82	0.30	0.42	0.28	0.92	0.26
2001	6.03	0.18	0.77	0.32	0.89	0.30	0.74	0.36
2002	4.07	0.19	0.70	0.26	0.94	0.19	0.40	0.34
2003	5.63	0.19	1.25	0.22	0.49	0.36	0.34	0.30
2004	7.17	0.18	2.24	0.18	1.06	0.28	0.48	0.25
2005	8.14	0.31	1.55	0.31	0.98	0.18	0.46	0.23
2006	5.30	0.20	1.39	0.35	0.97	0.33	0.40	0.22
2007	6.86	0.20	1.86	0.37	0.92	0.40	0.25	0.45
2008	20.15	0.61	2.07	0.26	0.87	0.24	0.92	0.28
2009	9.01	0.21						

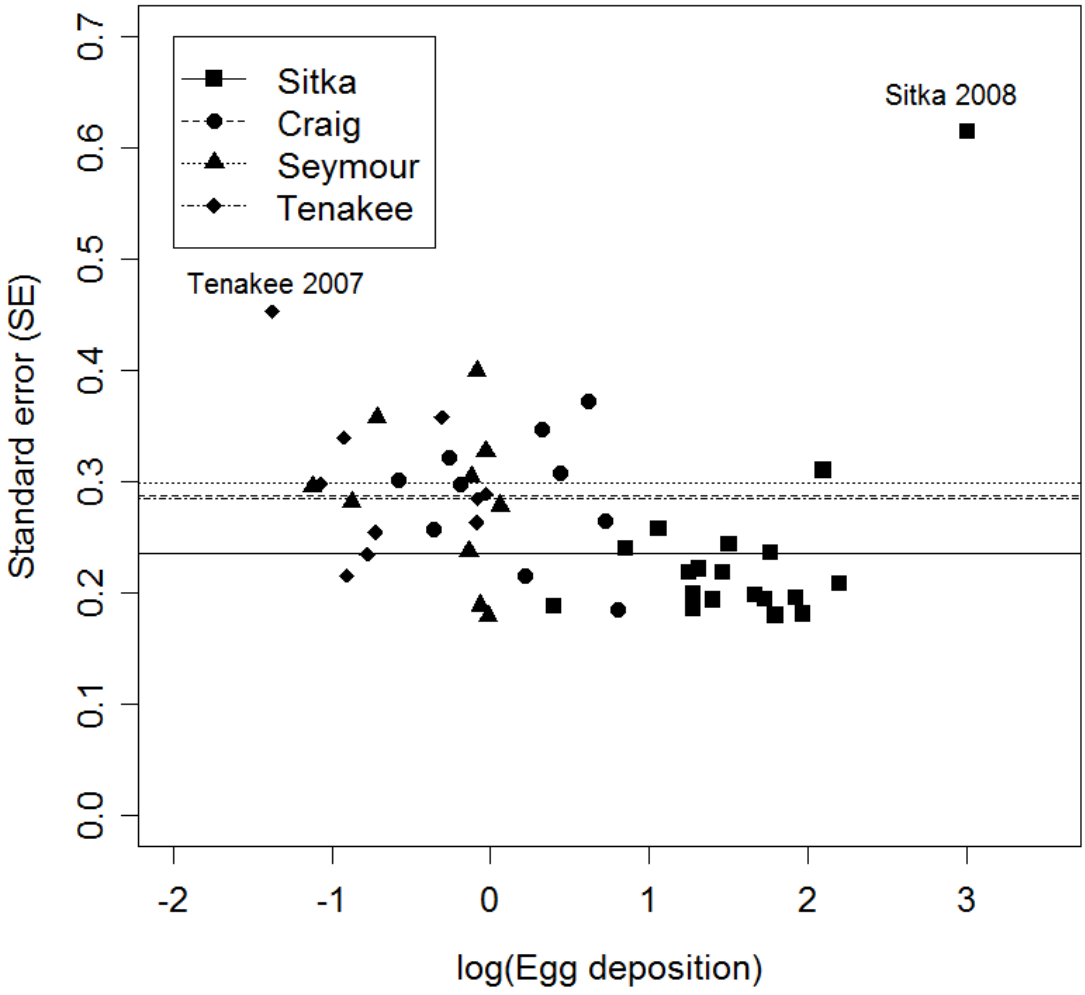

Figure 4.A.1. Log values of annual egg deposition for each herring population compared to the estimated standard error (SE). Obtained from resampling annual transect – quadrat data from the egg deposition survey conducted by ADF&G. Horizontal lines indicate the mean value for each population.

Appendix 4.B: Evaluation of the Dirichlet likelihood to fit age composition datasets

4.B.1 Introduction

Within this appendix, comparison is made between using the normal likelihood and the Dirichlet likelihood to fit age composition datasets in the individual stock age-structured assessment (ASA) models for the four Pacific herring stocks considered in this study that spawn in Sitka Sound, Craig, Seymour Canal, and Tenakee Inlet. The ultimate goal of this appendix is to determine which likelihood to employ in the metapopulation age-structured assessment (MPASA) model. The Alaska Department of Fish and Game (ADF&G) currently uses the normal likelihood to fit observations of age composition from the commercial fisheries and fishery-independent spawn surveys (Carlile *et al.* 1999). The Dirichlet likelihood, in which mean effective sample size (ESS) for a particular age or length composition dataset is estimated as a parameter, has been proposed to fit observations of age or length composition (Williams and Quinn 1998, Schnute and Haigh 2007, Chapter 2) and will be evaluated in this appendix.

4.B.2 Methods

Age composition datasets that are available for the four Pacific herring stocks considered in this study come from age composition of the commercial catch, and fishery-independent observations of age composition from the spawning population. Currently, age composition datasets are fit with the normal likelihood by ADF&G in the ASA models, given by:

$$-\ln L_{AC} = \lambda \sum_{t} \sum_{a} \left(\hat{\theta}_{a,t} - \theta_{a,t} \right)^2 \qquad (4.B.1)$$

in which $-\ln L_{AC}$ is the negative log-likelihood for the age composition datasets (*AC*), λ is the dataset weighting (proportional to the inverse of the variance, currently set at 1), $\hat{\theta}_{a,t}$ is the estimated age composition and $\theta_{a,t}$ is the observed age composition for age-*a* in time-*t*. The current methodology is a weighted least-squares scheme (e.g., Quinn and Deriso 1999). The Dirichlet likelihood for age composition is (following Chapter 2):

$$-\ln L_{AC} = \sum_{t} \left\{ \sum_{a} \left[\ln \Gamma\left(\hat{n}_{AC} \hat{\theta}_{a,t} \right) - \left(\hat{n}_{AC} \hat{\theta}_{a,t} - 1 \right) \ln\left(\theta_{a,t} \right) \right] - \ln \Gamma \hat{n}_{AC} \right\} \qquad (4.B.2)$$

where $\hat{n}_{AC}$ is the estimated mean ESS for the age composition dataset *AC* and is estimated as a parameter within the ASA model.

4.B.3 Results – Conclusions

Overall, trends in spawning biomass were similar between the individual ASA models that used the normal or Dirichlet likelihoods to fit age composition datasets (Figure 4.B.1). Spawning biomass from models that employ the normal likelihood for age

composition were outside of the 95% confidence intervals for spawning biomass from the Dirichlet likelihood in 1988 – 1993 for Craig, in 2002 and 2003 for Seymour canal, and in 1998 and 1999 for Tenakee. For Sitka, using the Dirichlet likelihood in the ASA model resulted in slightly larger recruitment estimates than using the normal likelihood (Figure 4.B.2). Recruitment estimates were smaller for the other populations from the Dirichlet likelihood compared to the normal likelihood, especially Tenakee in 1996, where the ASA model using the normal likelihood estimated recruitment to be upwards of 7,000 million fish.

Estimates of natural mortality were reduced when using the Dirichlet likelihood to fit age composition datasets, the largest being mortality estimates prior to 1998 (Figure 4.B.3). Further, the range in natural mortality among stocks was larger using the normal likelihood compared to the Dirichlet likelihood for estimates of natural mortality both prior to and after 1999. The estimated standard errors (SE) in natural mortality were greatly reduced when using the Dirichlet likelihood compared to the normal likelihood (Figure 4.B.3). Indeed, the SEs were so large using the normal likelihood that estimates of natural mortality for each stock were not significantly greater than 0, unlike the results from the Dirichlet likelihood (Figure 4.B.3). Additionally, parameter uncertainty was reduced for all estimated parameters using the Dirichlet likelihood compared to the normal likelihood (results not shown).

Estimates of mean ESS when using the Dirichlet likelihood for the catch age composition and spawning age composition ranged from 20-50 for both the individual stock ASA models (M0) and the preferred MPASA model (M2, Figure 4.B.4). Further,

the parameter estimates of mean ESS were well within the range of annual output ESS for both the ASA and MPASA models fit to the age composition datasets (Figure 4.B.4) indicating that the input and output uncertainty in age composition is similar.

Due to the reduced uncertainty in parameters estimated in the ASA models when using the Dirichlet likelihood to fit observations of age composition, we selected the Dirichlet likelihood rather than the normal likelihood to fit age composition datasets for further investigation of the MPASA model.

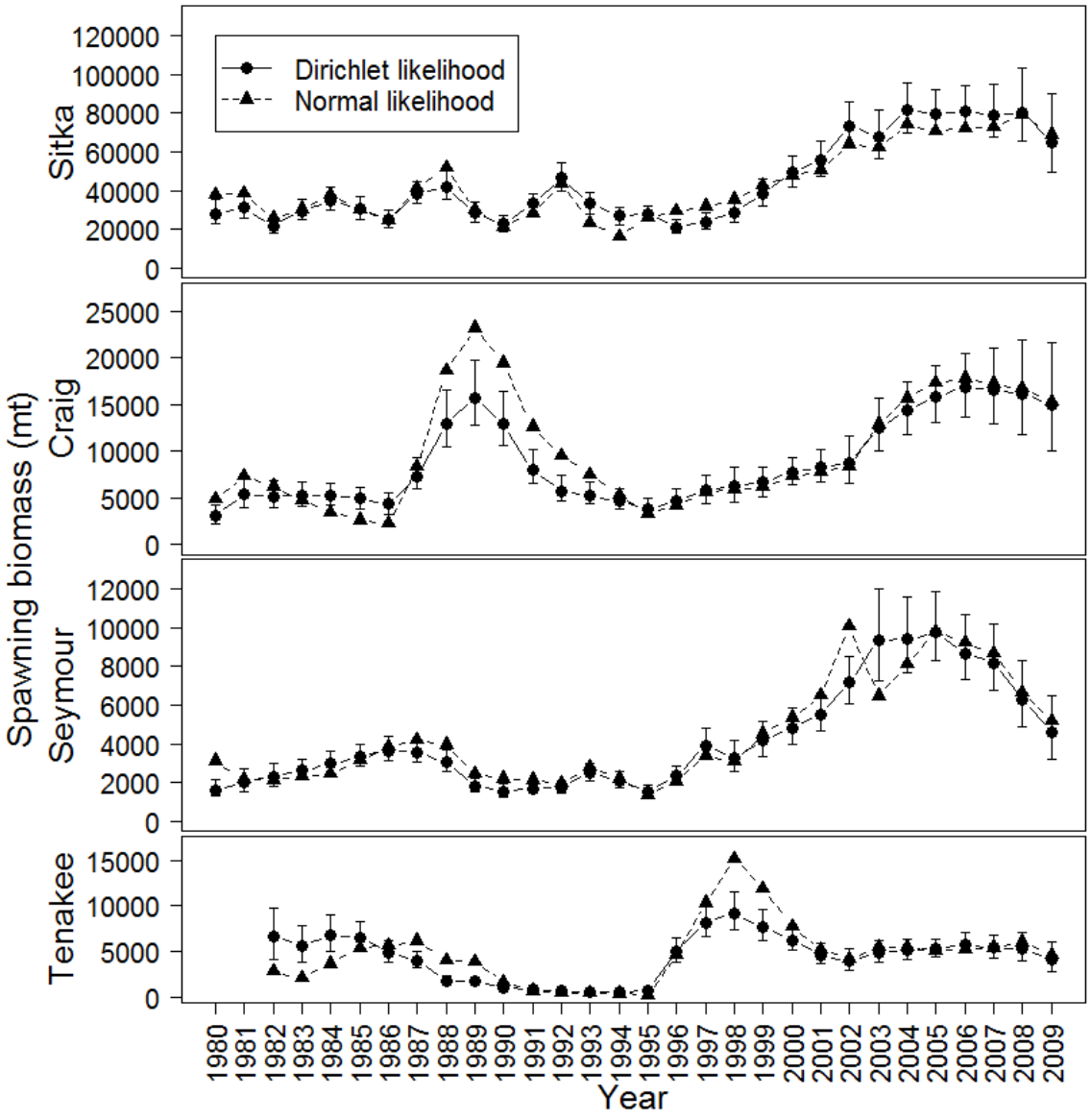

Figure 4.B.1. Individual stock ASA model estimated spawning biomass. Shown for models that fit age composition datasets with the Dirichlet likelihood or normal likelihood, 95% credible intervals from MCMC replication are shown with the Dirichlet likelihood estimated spawning biomass.

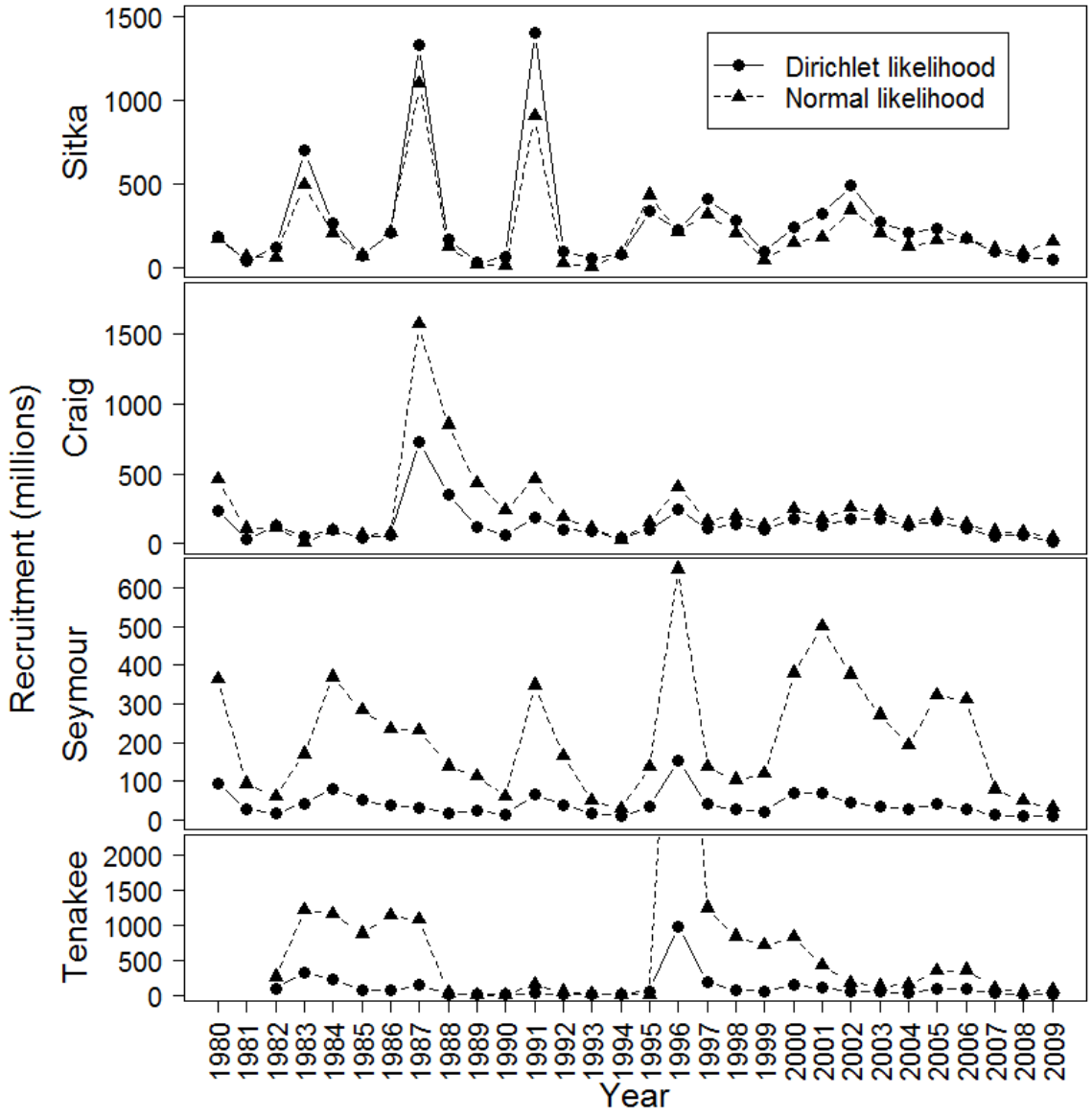

Figure 4.B.2. Individual stock ASA model estimated recruitment. Shown for models that fit age composition datasets with the Dirichlet likelihood or normal likelihood.

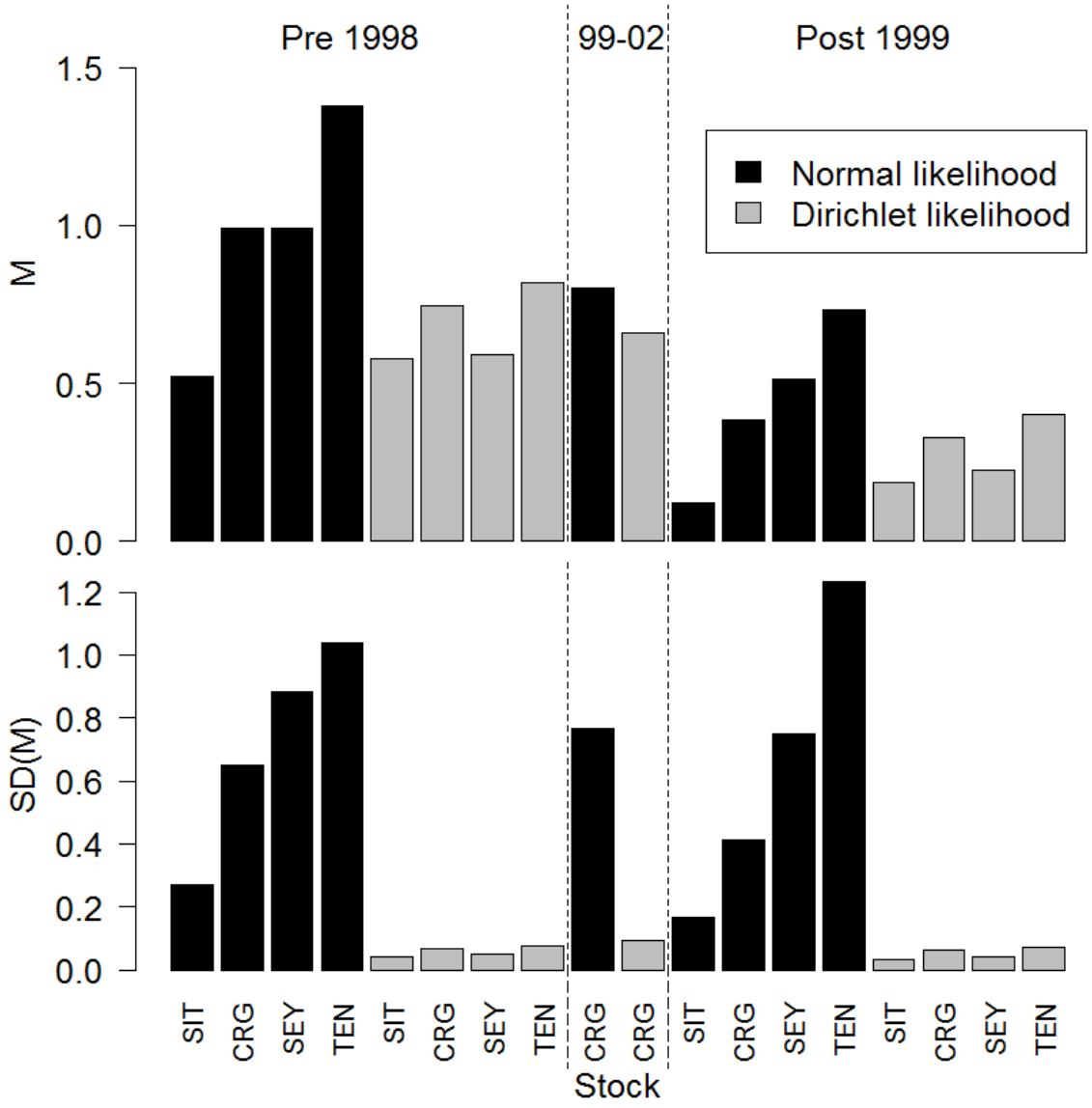

Figure 4.B.3. Estimated natural mortality with standard deviations. Individual stock ASA model estimated natural mortality parameters (top panel) with standard deviations (SD, bottom panel) for the four Pacific herring stocks considered (Sitka: SIT, Craig: CRG, Seymour: SEY, and Tenakee: TEN) that fit age composition datasets with the Dirichlet likelihood or normal likelihood.

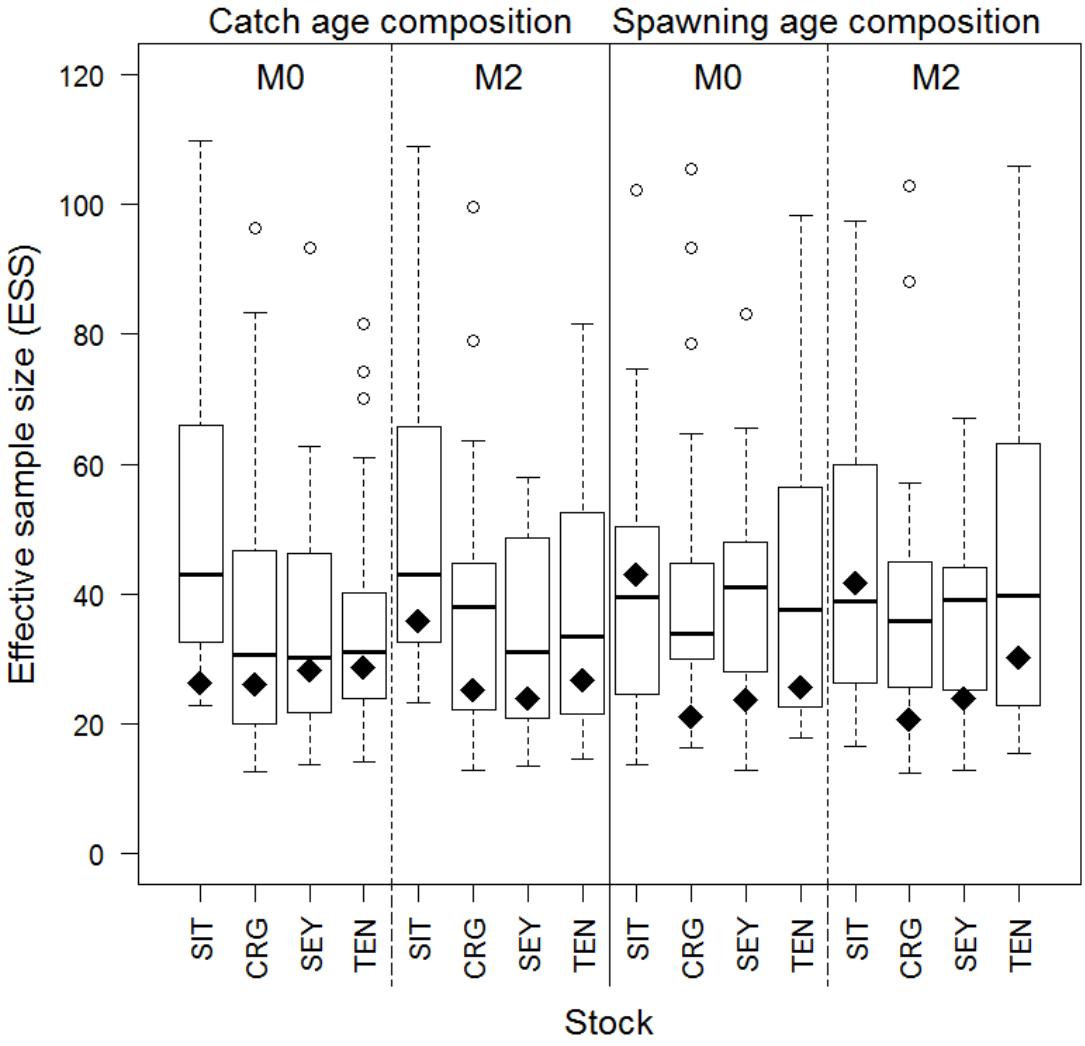

Figure 4.B.4. Estimated effective sample size (ESS) compared to output ESS. Parameter estimates of ESS are shown with diamond points and output ESS with boxplots for catch and spawning age composition of the four stocks considered (Sitka: SIT, Craig: CRG, Seymour: SEY, and Tenakee: TEN) in the individual ASA models (M0) and the preferred MPASA model (M2).

General Conclusions

A fundamental barrier to complete objectivity in integrated ASA models is the uncertainty specified for datasets in the likelihood function, which is equivalent to the choice of dataset weightings in the objective function. At present, the development of statistical theory has not provided objective methods to determine likelihood uncertainty for modeled datasets and estimated parameters in complex, integrated ASA models. Selection of weighting has been analyzed in integrated assessment models when supplemented by sample variance (Sullivan *et al.* 1999, Kimura 1989, 1990) or expert opinion (Merritt and Quinn 2000) and recent developments have investigated estimation of weighting within the integrated assessment model (McAllister and Ianelli 1997, Maunder and Starr 2003, Francis 2005, Deriso *et al.* 2007). Further studies have been devoted to evaluating methods to overcome this obstacle (Schnute and Richards 1995, Fournier *et al.* 1998, Methot 2000, Wilberg and Bence 2006). In addition, studies have examined the effectiveness of the integrated assessment model with uncertainty in data (Booth and Quinn 2006), its subsequent effect on model conclusions (Deriso *et al.* 2007, Hulson *et al.* 2008) and risk to fisheries management (Hilborn *et al.* 1993, McAllister and Ianelli 1997). An objective method to determine dataset weighting is of critical interest to all assessment scientists, as it can influence not just model predictions but the uncertainty in model predictions as well.

Effective sample size (ESS) is a quantity that accommodates unknown process or model misspecification error in age and length composition data (e.g., McAllister and Ianelli 1997). By investigating ESS in chapter one, I was able to isolate and evaluate

possible sources of error in age and length composition datasets and compare between multiple methods of approximating ESS. The sampling approximation method resulted in large variability in ESS, although, ESS can be approximated directly from survey and fishery observations of age and length composition (Pennington *et al.* 2002). Using the Dirichlet likelihood to approximate ESS though maximum likelihood procedures (e.g., Williams and Quinn 1998) resulted in smaller standard deviations in ESS compared to the variance ratio method (McAllister and Ianelli 1997), although, both methods require knowledge of the true age or length composition of the population studied, which is rarely, if ever, known in practical applications.

For age and length composition datasets integrated in ASA models I found several sources of error that can affect the uncertainty of samples obtained from surveys or fisheries in chapter one. In terms of model specification error, the greatest influence on the uncertainty of age and length composition datasets was the age aggregation of fish within schools. Uncertainty in age and length composition increased as the schools from which samples were obtained contained fewer age classes. In terms of sampling design, uncertainty can be decreased by obtaining samples of fish from a greater number of schools. From chapter one, the recommendation for obtaining samples for age and length composition is to sample from a larger number of schools. If there are constraints on the total number of samples that can be obtained due to cost, for example, uncertainty can be reduced by taking a smaller number of samples from any given haul while keeping the total number of samples collected the same.

From the results of chapter two, I found that it may be possible to estimate the uncertainty of age and length composition datasets though ESS within an ASA model. By using the Dirichlet distribution and treating the number of samples as a random variable, the effective sample size can be parameterized. I found that estimating mean effective sample size across years rather than estimating annual effective sample size was more parsimonious; estimation of annual deviations in effective sample size is a topic of further investigation. Compared to historical methods to incorporate effective sample size, use of the Dirichlet distribution reduced bias and uncertainty in model predictions, although there were a number of historical methods that resulted in accurate estimation of management quantities. A limitation of ASA models was also discovered in chapter two. When samples for age and length composition were obtained from schools that were composed of a single age class, not a single method to incorporate effective sample size resulted in accurate parameters or management quantities. From the results of chapter two, I recommend that further investigation be focused on using the Dirichlet distribution to fit age and length composition and estimate effective sample size in ASA models used for stock assessment; the uncertainty in model predictions would then incorporate uncertainty in the datasets through parameterization of variance.

In chapter three, I found that if climate change results in a latitudinal shift of distribution of fish populations, spatially explicit ASA models would provide accurate results of biomass and management quantities. I also found that ASA models that did not consider spatial information also provide accurate management quantities, but the uncertainty is larger than if a spatially explicit ASA model were used. In the case of

parameter misspecification, especially ontogenetic movement, spatially explicit ASA models resulted in bias for biomass. When the functional form of the effect of climate change on movement parameters is unknown, a random walk can be employed to aid in accurate estimation of biomass and management quantities. This could also be done for other parameters, such as natural mortality, selectivity, and catchability. Based on the results of chapter three I recommend that efforts be made to employ spatially-explicit ASA models for management and that random walks be employed to evaluate annual deviations in parameters, as the uncertainty in model predictions is smaller than if an ASA model was used that did not consider spatial information.

In chapter four, I constructed a metapopulation ASA model for Pacific herring in southeast Alaska. The metapopulation ASA model resulted in smaller uncertainty in model parameters and population quantities than the individual stock ASA models. I also found that through sharing time dependent parameter deviations in natural mortality and maturity there may be localized effects on natural maturity coupled with more broad scale effects on maturity. These localized effects may be due to differing oceanographic domains in the inside passages of southeast Alaska and the regions open to the Gulf of Alaska (e.g., Williams and Quinn 2000). From the results of chapter four, I recommend that the metapopulation ASA model for herring stocks in southeast Alaska be extended and used for management due to the reduced uncertainty in model output.

An intermediate step in moving towards ecosystem based fisheries management (EBFM) is to improve current stock assessment models by developing spatially explicit (Quinn and Collie 2005, Goethel *et al.* 2011) and metapopulation stock assessment

models (Kritzer and Sale 2004). In terms of making EBFM operational, spatially explicit models would evaluate interactions among aquatic species at finer spatial scales, as modeling at broad geographic scales can misinterpret fundamental population dynamics (Härkönen and Harding 2001). Further, metapopulation models would allow investigations into both broad and localized scale influences on population dynamics for a number of fish stocks simultaneously (Smedbol *et al.* 2002). Development of EBFM will be accomplished through understanding of fish populations at finer spatial scales and with more realistic stock structures (i.e., metapopulation structure). Compared to historical ASA models, both the spatially explicit and metapopulation ASA models reduced uncertainty in model results and resulted in accurate estimates of population parameters and quantities even if parameters were altered over time due to climate change.

The Deviance Information Criterion (DIC) and the evaluation of effective number of parameters were additional tools to evaluate ASA model uncertainty. Essentially, DIC is a measure of how well the ASA model predicts the data that is used to estimate population parameters (Spiegelhalter *et al.* 2002). The effective number of parameters is also a measure of model uncertainty and can be used to compare to the actual number of parameters in an ASA model (Wilberg and Bence 2008). By comparing the likelihood value over Markov-chain Monte Carlo (MCMC) replicates to the likelihood value when the predictions equal the observed value, one obtains a measure of model uncertainty, the main method used to compute both the DIC value and effective number of parameters of an ASA model. Through using DIC, I was able to evaluate parameterizations within the

spatially explicit and metapopulation ASA models to evaluate both model comparison and model uncertainty. In chapter three, I also found that across replicated datasets the DIC value can be variable, and that model comparison can favor several parameterizations based on the error in datasets used in an ASA model. Using DIC and effective number of parameters in chapter four also aided in evaluation of metapopulation scale model scenarios.

Two main perspectives to deal with uncertainty in ASA models were presented in this thesis. The first perspective focused on uncertainty in datasets integrated in ASA models and how this uncertainty relates to model output. From the results presented herein, there may be a way to estimate dataset uncertainty within ASA models, possibly resolving the issue of stock assessment scientists choosing relative uncertainty between integrated datasets in ASA models. By using ESS and the Dirichlet likelihood, stock assessment scientists can deal with unknown sources of dataset uncertainty in age and length composition and estimate this uncertainty in the ASA model. The second perspective focused on model structure. Use of spatially-explicit ASA models and metapopulation ASA models indeed provided improvement over non-spatial ASA models and individual stock ASA models. However, the approaches used to evaluate uncertainty in ASA models in this thesis are not exhaustive.

As stock assessment models become increasingly complex the questions of uncertainty will also become more complex. The primary tools evaluated herein to understand model uncertainty included using ESS with age and length composition data and evaluating DIC and the effective number of parameters (e.g., Spiegelhalter *et al.*

2002, Wilberg and Bence 2008). Estimating ESS will allow for more comprehensive evaluation of DIC and effective number of parameters, as a parameter related to data uncertainty would also include uncertainty in the MCMC replication of model parameters. This approach also allows for evaluation of model uncertainty while taking into consideration dataset uncertainty. While all the questions regarding uncertainty in ASA models have yet to be resolved, this thesis provides guidance for further evaluation of uncertainty in ASA models through simulation and evaluation of ASA model structure.

References

Allen, M. J., and Smith, G. B. 1988. Atlas and zoogeography of common fishes in the Bering Sea and northeastern Pacific. U.S. Department of Commerce, NOAA Technical Report NMFS 66.

Booth, A. J., and Quinn, T. J., II, 2006. Maximum likelihood and Bayesian approaches to stock assessment when data are questionable. Fisheries Research, 80: 169–181.

Caswell, H. 2000. Matrix Population Models: Construction, Analysis and Interpretation, 2nd ed., Sinauer Associates, Sunderland, MA.

Deriso, R. B., Quinn, T. J., II, and Neal, P. R. 1985. Catch-age analysis with auxiliary information. Canadian Journal of Fisheries and Aquatic Sciences, 42: 815–824.

Deriso, R. B., Maunder, M. N., and Skalski, J. R. 2007. Variance estimation in integrated assessment models and its importance for hypothesis testing. Canadian Journal of Fisheries and Aquatic Sciences, 64: 187–197.

Doubleday, W. G. 1976. A least squares approach to analyzing catch at age data. International Commission of Northwest Atlantic Fisheries Research Bulletin, 12: 69–81.

Fournier, D., and Archibald, C. P. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39: 1195–1207.

Fournier, D. A., Sibert, J. R., Majkowski, J., and Hampton, J. 1990. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency datasets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). Canadian Journal of Fisheries and Aquatic Sciences, 47: 310–317.

Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunge*. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105–2116.

Francis, R. I. C. C. 2005. Assessment of hoki (*Macruronus novaezelandiae*) in 2004. N.Z. Fishery Assessment Report No. 2005/35.

Fry, F. E. J. 1949. Statistics of a lake trout fishery. Biometrics, 5: 27–67.

Fu, C., and Quinn, T. J., II, 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Canadian Journal of Fisheries and Aquatic Sciences, 57: 2420–2432.

Gulland, J. A. 1965. Estimation of mortality rates. Annex to Arctic fisheries working group report. ICES C.M. Document No. 3. (Mimeograph).

Goethel, D. R., Quinn, T. J., II, and Cadrin, S. X. 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Reviews in Fisheries Science, 19(2): 119–136.

Hanselman, D. H., Heifetz, J., Fujioka, J., and Ianelli, J. N. 2003. Gulf of Alaska Pacific ocean perch. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2004. North Pacific Fishery Management Council, Anchorage, AK.

Hanselman, D. H., Heifetz, J., Fujioka, J., Shotwell, S. K., and Ianelli, J. N. 2007. Gulf of Alaska Pacific ocean perch. *In* Stock Assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2008. North Pacific Management Council, Anchorage, AK.

Härkönen, T., and Harding, K.C. 2001. Spatial structure of harbour seal populations and the implications thereof. Canadian Journal of Zoology, 79: 2115–2127.

Hay, D. E. 1985. Reproductive biology of Pacific herring (*Clupea harengus pallasi*). Canadian Journal of Fisheries and Aquatic Sciences, 47: 2390–2401.

Hilborn, R., Pikitch, E. K., and Francis, R. C. 1993. Current trends in including risk and uncertainty in stock assessment and harvest decisions. Canadian Journal of Fisheries and Aquatic Sciences, 50: 874–880.

Hulson, P.-J. F., Miller, S. E., Quinn, T. J., II, Marty, G. D., Moffit, S. D., and Funk, F. 2008. Data conflicts in fishery models: incorporating hydroacoustic data into the Prince William Sound Pacific herring assessment model. ICES Journal of Marine Science, 65: 25–43.

Ianelli, J. N., Barbeaux, S., Honkalehto, T., Kotwicki, S., Aydin, K., and Williamson, N. 2007. Eastern Bering Sea Walleye Pollock. *In* Stock Assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands as projected for 2008. North Pacific Management Council, Anchorage, AK.

Jones, R. 1964. Estimating population size from commercial statistics when fishing mortality varies with age. ICES Journal of Marine Science, 155(38): 210–214.

Kimura, D. K. 1989. Variability, tuning, and simulation for the Doubleday-Deriso catch-at-age-model. Canadian Journal of Fisheries and Aquatic Sciences, 46: 941–949.

Kimura, D. K. 1990. Approaches to age-structured seperable sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences, 47: 2364–2374.

Kritzer, J. P., and Sale, P. F. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries, 5: 131–140.

Marty, G. D., Quinn, T. J., II, Carpenter, G., Meyers, T. R., and Willits, N. H. 2003. Role of disease in abundance of a Pacific herring population. Canadian Journal of Fisheries and Aquatic Sciences, 60: 1258–1265.

Maunder, M. N., and Starr, P. J. 2003. Fitting fisheries models to standardized CPUE abundance indices. Fisheries Research, 63: 43–50.

McAllister, M. K., and Ianelli, J. N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54: 284–300.

Merritt, M. F., and Quinn, T. J., II, 2000. Using perceptions of data accuracy and empirical weighting of information: assessment of a recreational fish population. Canadian Journal of Fisheries and Aquatic Sciences, 57: 1459–1469.

Methot, R. D. 2000. Technical description of the stock synthesis assessment program. NOAA Technical Memorandum NMFS-NWFSC-43.

National Research Council (NRC). 1998. Improving Fish Stock Assessments. National Academy Press, Washington, DC.

Pennington, M., Burmeister, L. M., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin US, 100: 74–80.

Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Research Bulletin for the International Commission of Northwest Atlantic Fisheries, 9: 65–74.

Quinn, T. J., II, and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Quinn, T. J., II, 2003. Ruminations on the development and future of population dynamics models in fisheries. Natural Resource Modeling, 16(4): 341–392.

Quinn, T. J., II, and Collie, J. S. 2005. Sustainability in single-species population models. Philosophical Transactions of the Royal Society B, 360: 147–162.

Schnute, J.T., and Richards, L. J. 1995. The influence of error on population estimates from catch-age models. Canadian Journal of Fisheries and Aquatic Sciences, 52: 2063–2077.

Smedbol, R. K., A. McPherson, M. M. Hansen, and E. Kenchington, 2002. Myths and moderation in marine 'metapopulations'? Fish and Fisheries, 3: 20–35.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, 64(4): 583–639.

Sullivan, P. J., Parma, A. M., and Clark, W. J. 1999. The Pacific halibut stock assessment of 1997. International Pacific Halibut Commission Scientific Report No. 79.

Wilberg, M. J., and Bence, J. R. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences, 63: 2275–2285.

Wilberg, M. J., and Bence, J. R. 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. Fisheries Research, 93: 212–221.

Williams, E. H., and Quinn, T. J., II, 1998. A parametric bootstrap of catch-age compositions using the Dirichlet distribution. *In* Fishery Stock Assessment Models, pp. 371–384. Ed. by F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, *et al.* Alaska Sea Grant Program Report AK-SG-98-01, University of Alaska, Fairbanks. 1054 pp.

Williams, E. H., and Quinn, T. J., II, 2000. Pacific herring, *Clupea pallasi*, recruitment in the Bering Sea and north-east Pacific Ocean, II: relationships to environmental variables and implications for forecasting. Fisheries Oceanography, 9(4):300–315.

Appendix A: Including mark-recapture data into a spatial age-structured model: walleye pollock (*Theragra chalcogramma*) in the eastern Bering Sea

Abstract

Integrated assessment models used to evaluate fish stocks are becoming increasingly complex; some are capable of incorporating spatial considerations. Such a model has been developed to estimate movement of walleye pollock (*Theragra chalcogramma*) between the northwest and southeast eastern Bering Sea. In this study, we investigate the feasibility of estimating movement using spatially disaggregated data supplemented by tagging data. Monte Carlo simulation was used to test accuracy and variability of parameter estimation in model scenarios with and without tagging information. Total biomass estimates for models with and without tagging data were unbiased, but uncertainty was smaller when tagging data were available. Uncertainty was also reduced in regional biomass and movement parameters when including tagging data. Our findings indicate that tagging information would be important to provide reliable spatially explicit fisheries management advice for eastern Bering Sea pollock.

A.1 Introduction

Because the true mechanisms and processes underlying the natural world are unknown, researchers describe the system using models based on available observations. These datasets typically lack the contrast needed to estimate more complex parameterizations because parameters can become confounded due to the statistical models employed. For example, natural mortality parameters are often confounded with movement parameters in spatial models, a case in which tagging data have been proposed as a way to address parameter confounding (Anganuzzi 1996).

One of the most powerful tools available to test assessment models in fisheries science is simulation (Hilborn and Mangel 1997). Monte Carlo simulation has been used to determine model robustness from incorrectly specifying the error structure or parameter structure (e.g., Fu and Quinn 2000). By using simulated data, the researcher knows the true parameter values and can then evaluate model performance (Hilborn and Mangel 1997). Through realistic simulation that includes aspects of random variability in dynamics and observation, it is possible to investigate the biases, effectiveness, and impacts on management procedures of new stock assessment methods (Hilborn and Walters 1992, Maunder 2001). This study presents an application of simulation-estimation tests for a relatively complex spatially-explicit movement model.

Changes to a fish population structure over space or time can affect fish stock assessment estimates of potential fishery yields and mortality. Assuming there is no movement in a fish population when movement actually occurs can lead to a loss in potential harvests (Tuck and Possingham 1994). Conversely, assuming movement when

populations are relatively stationary over space or time can lead to depletion imbalances with fishing (Hilborn and Walters 1992). Incorporation of movement in fish stock assessment analysis is rare due to lack of tagging information and added model complexity (Quinn and Deriso 1999). Consequently, spatial assessment models are uncommon in fishery management settings but there have been some promising developments.

An age-structured movement model has been developed for eastern Bering Sea (EBS) walleye pollock (*Theragra chalcogramma*, hereafter referred to as pollock) which estimates movement based on spatially disaggregated survey and fishery data (Miller et al. 2008). A method for including tagging data into assessment models was provided by Maunder (2001). In this model the tagged fish follow the same population dynamics as the entire population and are essentially modeled as a subset of the entire population within the assessment model. Other spatial models that include tagging data include that of Punt et al. (2000) and Hanselman et al. (2007). The first is a spatially-explicit model for school shark (*Galeorhinus galeus*) off southern Australia that includes tagging data. The latter is a size-based migration model designed for the migratory Alaska sablefish (*Anaplopoma fimbria*) that uses mark-recapture datasets that began in 1979 (Heifetz and Fujioka 1991). Also, a size-structured model that includes mark-recapture information for South Australian rock lobster has been developed (McGarvy et al. 2010) and is used for management advice.

The groundfish fishery for pollock in the EBS is one of the world's largest, and catch limits are managed by seasons. During spring and summer in the EBS, pollock are

thought to migrate to feeding areas, and during the late fall and early winter they migrate to spawning areas. Pollock distribution with respect to temperature obtained from EBS surveys shows that feeding migrations tend to follow a northward and shoreward movement, the extent of which increases with increasing size of pollock (Kotwicki et al. 2005). Currently there is no movement information from a large-scale EBS pollock mark-recapture study, but feasibility studies for tagging of pollock have shown the potential for sufficient sample size and survival rates (Natural Resources Consultants, Inc. and Northwest Marine Technology, Inc. 1996, Miller 2007, Winter et al. 2007).

Using disaggregated survey and fishery catch data, age-specific pollock movement between the northwest (NW) and southeast (SE) EBS were estimated using an age-structured assessment movement model (ASM, Miller et al. 2008). This model differs from the present EBS pollock stock assessment that is based on a standard age-structured stock assessment model (Ianelli et al. 2009) and does not estimate movement. Under moderate assumptions for spatial structure (the number of regions in which movement was estimated) reasonable estimates of most population and movement parameters could be obtained. However, some parameter estimates were highly uncertain and correlated, and the analysis lacked a formal evaluation of accuracy and precision. In the current study, the utility of additional information from a hypothetical mark-recapture program is examined and compared to the ASM model constructed by Miller et al. (2008). We do not consider inter-population movement; rather we assume that the pollock ‘stock’ exhibits intra-population movement between the two regions in the EBS. The ability to reliably estimate movement parameters was tested using Monte Carlo

simulation (Hilborn and Mangel 1997). Given specified spatial and seasonal population characteristics, we investigated the bias and variability in parameters and population estimates across a variety of scenarios including the addition of tagging data. The goals were to determine: (1) the benefits of having tagging datasets for the ASM model, (2) the sensitivity of results to tag release protocols (e.g., season and area of releases), and (3) the magnitude of bias and variability in population parameter estimates without tagging data.

A.2 Material and Methods

In this study the ASM model developed by Miller et al. (2008) was used as the operating model. Estimates based on the original data fit with the ASM model formed the 'true' state of nature from which data were generated (with specified errors) for subsequent fitting by the estimation models (Table A.1). The two regions considered for the Eastern Bering Sea in Miller at al. (2008) were divided into NW and SE by the 170° W meridian (Fig. A.1). The datasets used in the ASM model are described in Miller et al. (2008). The Aleutian Islands and the Gulf of Alaska were not included in the analysis.

As the primary focus of this study was to evaluate the effect of tagging data on ASM model estimates, tagging data by region and season were also generated and applied to an Integrated Tagging and Catch-at-Age Analysis **(**ITCAAN) estimation model (Maunder 2001). Results from the simulation analyses were compared through bias and variability in the movement parameters and biomass estimates. The simulation procedure was executed with customized code in R language (R Development Core Team 2008)

that integrated R with AD model builder (ADMB, Fournier 1996). The ASM and ITCAAN estimating models were written using ADMB and the operating model and simulation-estimation loops were written using R. For each variance scenario, 1000 operating model simulations were evaluated; the number of simulations was selected to obtain robust estimates of variance (e.g., Efron and Tibshirani 1993). Further, the ASM and ITCAAN models in each simulation used the same generated data except the generated tag recovery data because the ASM model does not fit tag recovery data.

A.2.1 Operating model

In Miller et al. (2008), the population dynamics of EBS pollock was modeled with a discrete-time and region-specific ASM model (Quinn and Deriso 1999, Maunder 2001). The generalized dynamics of pollock in region *r* at age *a* in year *y* at the start of the year in the ASM model can be formulated as the sequence of spatial (NW, SE) and temporal (A and B seasons) events shown in Figure A.2. Pollock spawn in the winter and this time period is used in the ASM model as the start of the modeling year (January 1st). The "A" season fishery occurs subsequent to the start of the year from January to June. After the "A" season, the first regional movement between the NW and SE takes place to reflect movement of pollock from the spawning to summer feeding grounds. During this time half of the year's natural mortality was applied to the population. The "B" season fishery is subsequent to the first regional movement and occurs from July to December. After the "B" season fishery, the second regional movement was modeled to reflect movement back to spawning grounds in the fall, and the second half of the year's natural mortality is

applied. Note that when movement occurs it is not the entire population moving from region to region, rather a proportion of the population that moves as determined by the probability of movement estimated by movement parameters. The population after the "A" fishing season at age *a* in year *y* was the first step (1) in the yearly population dynamics equation:

$$ {}^{1}_{r}N_{a,y} = {}_{r}N_{a,y}e^{-{}_{r}F_{a,y,A}} \tag{A.1} $$

Movement from the winter spawning grounds to summer feeding grounds (step 2), with half-year natural mortality, was modeled as:

$$ {}^{2}_{r}N_{a,y} = \left({}^{1}_{r}N_{a,y}\theta_{r\rightarrow r,A} + {}^{1}_{r'}N_{a,y}\theta_{r'\rightarrow r,A}\right)e^{-M/2} \tag{A.2} $$

in which the abundance in region *r* is the sum of the numbers of fish that stay in region *r* ($r \rightarrow r$) and the numbers that leave region *r'* for region *r* ($r' \rightarrow r$). The "B" season fishing mortality was applied to the population in step 3:

$$ {}^{3}_{r}N_{a,y} = {}^{2}_{r}N_{a,y}e^{-{}_{r}F_{a,y,B}} \tag{A.3} $$

After the "B" fishing season movement from the summer feeding grounds back to the spawning grounds, the final step in the yearly population dynamics (4), was estimated by:

$$ {}^{4}_{r}N_{a,y} = \left({}^{3}_{r}N_{a,y}\theta_{r\to r,B} + {}^{3}_{r'}N_{a,y}\theta_{r'\to r,B} \right) e^{-M/2} \qquad \text{(A.4)} $$

which becomes the population abundance at the beginning of the next year. The prime notation is used to describe the alternate region, that is, if r = NW then r' = SE. In these equations F represents fishing mortalities for the A and B season, M is natural mortality, and θ is the probability of movement after fishing seasons A or B from region r to region r', which may also be a function of age.

Pollock are thought to move seasonally and with a general ontogenetic pattern that shifts southeastward with increasing age (Buckley et al. 2001). For the NW region, the operating model was conditioned such that the probability of movement from the NW to SE was an increasing linear function of age. For convenience the modeling of movement was parameterized in terms of probability of staying in the same region (1 – probability of movement). Based on available knowledge, the probability that age 3 fish stayed in the NW after the A season was $\theta_{\mathrm{NW}\to\mathrm{NW,\,A}}$ = 99% and the probability after the B season was $\theta_{\mathrm{NW}\to\mathrm{NW,\,B}}$ = 75%. These amounts were multiplied by 80% for each successive age class for movement after the A fishing season and by 90% for the movement after the B fishing season (Miller et al. 2008). For the SE EBS, available knowledge suggests that the probability of staying in the SE region was constant for all ages, equal to $\theta_{\mathrm{SE}\to\mathrm{SE,\,A}}$ = 30% after the A season and $\theta_{\mathrm{SE}\to\mathrm{SE,\,B}}$ = 70% after the B season (Miller et al. 2008). These four parameters were estimated in the estimation model. The corresponding age-specific movement rates are shown in Figure A.3.

A.2.2 Data Generation

Data were simulated by generating random variability about the expected values from the operating model and fitting the generated data with the estimation model structures. In the simulation analysis, a total of 13 datasets given in Table A.2 were generated and employed in the ASM model structure and 17 in the ITCAAN model structure. The four additional datasets in the ITCAAN model were tag return data from the two regions and two fishery seasons (Table A.3). Because the log-normal distribution was used in the ASM model fit by Miller et al. (2008), simulated data were structured to follow a lognormal distribution to provide a direct comparison with the prior study:

$$x_i = \bar{x} e^{\left(\varepsilon_i - \frac{\sigma^2}{2}\right)} \qquad \varepsilon_i \sim N(0, \sigma) \tag{A.5}$$

in which $\bar{x}$ is the "true" value from an area and year, ε_i is the randomly generated normal deviation for simulation *i*, and σ is the pre-specified standard deviation. The values for σ in the estimation model were set to a range of different values relative to the 'true' operating model values (Table A.2 and Table A.3).

Because tagging information was the key focus in this study, 'observed' tag recovery data were constructed in the operating model by a simulated tagging program. In this program, there were annual releases of tags into the marked population in both regions for all ages 3 to 10+ during the summer surveys. To evaluate the effect of the

number of years that tag release and recovery were conducted, two cases were considered. In the first case (I) tagged fish were released in each year from 2000 to 2002. In the second case (II) tagged fish were released each from 1995 to 2002. Recoveries for both cases start in the first year of tagging and continue through 2004. We refer to a spatial model with both regular assessment data and mark-recapture data as an Integrated Tagging and Catch-at-Age Analysis **(**ITCAAN) model after Maunder (2001).

In the ITCAAN model, the tagged and the total population (including both tagged and untagged fish) were modeled simultaneously with the same dynamics and parameter values (e.g., Maunder 2001), except that the initial tagged populations at the start of the mark-recapture program were the numbers of releases by regions and ages. It was assumed that the pollock would be implanted with magnetically-detectable coded wire tags (CWT) by a research vessel in the summer during the bottom trawl survey (BTS), such that fish in different regions would receive different tags. In this study we did not consider tag induced mortality, as an actual tagging program has not been implemented for pollock and it would be difficult to presume what tag induced mortality would be. It was further assumed that fish would be recovered by the commercial fishery in either the A or B fishery season. There are three fishing sectors in the EBS: (1) fishing vessels with shoreside processors (50% of Total Acceptable Catch, TAC), (2) fishing vessels with offshore mothership processors (10% of TAC), and (3) vessels that both harvest and process their catch offshore (40% of TAC). Only the feasibility of installing electronic CWT detectors in shoreside processing plants has been determined (NRC 1996).

Therefore, only the proportion of the catch typically caught by the fishing vessels with shoreside processors was considered in the tag recovery process.

The number of recoveries ($^{r}R_{a,y,s}$) at age *a* in year *y* from region *r* and season *s* was generated by:

$$^{r}R_{a,y,s} = {}^{r}p \; {}^{r}T_{a,y,s}\left(1 - e^{-{}^{r}F_{a,y,s}}\right) \qquad \text{(A.6)}$$

in which $^{r}T_{a,y,s}$ is the number of marked individuals at age *a* in year *y* from region *r* and season *s* after half a year of natural mortality, *F* is the fishing mortality, and ^{r}p is the proportion of the total catch delivered to shoreside processors from region *r*. For simplicity, this proportion was treated as constant over time, which is reasonable from examination of catch data. Recoveries are a function of the exploitation fraction of the fishery; incidental recoveries from survey effort were omitted since total catch during a survey is very low. The proportion of total catch from region *r* (^{r}p) delivered to shoreside processors, derived from observed regional landings from 2003-2005, was set at 7% from the NW region, and 39% from the SE region (Miller 2007). As an actual tagging program has not been implemented for pollock, in this simulation study it was assumed that there was no tag loss or tag induced mortality for fish that were tagged with CWTs and that the total catch delivered to shoreside processors was examined for tags with a detection rate of 100%.

Two methods of releasing tags during the survey were examined in the simulation analysis. In the first method, tags were released evenly across the NW and SE region. In

this case 5,000 tags were released in the NW and 5,000 in the SE during the BTS survey. In the second method tags were released proportionally to the BTS survey catch: 67% (6,700) of the tags were released in the NW region and 33% (3,300) were released in the SE during the survey. In both cases it was assumed that the ages at the time of tagging were known, the tags were distributed uniformly across age classes, and the tagged fish mixed completely in the population with untagged fish. Fitting the ITCAAN model to even tagging is denoted by $ITCAAN_E$, and fitting to proportional tagging is denoted $ITCAAN_P$. The two cases for the number of years for which tags were released for each tagging method were denoted as (evenly tagging, for example) $ITCAAN_{E,I}$ for case I and $ITCAAN_{E,II}$ for case II.

Two uncertainty scenarios for the tag recovery data were used in the simulation (Table A.3). In the first scenario (a), the standard deviation for tag recoveries was set at 0.05 to represent a case in which the uncertainty in tag returns was low. The second scenario (b) used standard deviations that were proportional to the inverse of the number of tag recoveries. This scenario was selected to evaluate uncertainty that is related to the sample size of tag recoveries obtained.

A.2.3 Estimation models

In both the ASM and ITCAAN models, parameter estimates were obtained using an objective function that represented the negative log likelihood of normalized residuals. The objective function consisted of 13 dataset sources for the ASM model in Table A.2 (Miller et al. 2008). For the ITCAAN model, there were the 13 dataset sources fitted in

the ASM with an additional 4 datasets for tag recoveries by fishing season and region Table A.3). The weighting terms, or equivalently the dataset standard deviations, were derived from either the 2009 stock assessment report (Ianelli et al. 2009) or were assigned a value consistent with available knowledge (Table A.2).

In the ITCAAN estimation model, the tag release data were directly input into the model, because this information would be known. The tag returns datasets were fit in the estimation model, after being estimated following equation A.6, with a lognormal likelihood. We only considered region specific tags, rather than also considering year and region specific tags; more sophisticated treatments of tagging data would potentially improve the ITCAAN model but are beyond the scope of this study. The lognormal rather than discrete distributions (i.e., Poisson, multinomial) was used in order to provide more direct comparison with Miller et al. (2008), as the lognormal likelihood was used in that study. Further analysis is being conducted that uses discrete distributions to generate tagging datasets. In each tag recovery dataset uncertainty scenario, the standard deviations used in data simulation were subsequently used in the objective function as the likelihood variance to fit the generated datasets. We did not consider misspecification of variance in the estimation model.

For presentation we focused on estimates of movement parameters and total and regional biomass in the results. For total and regional biomass the results focus on the estimate in the last year of the model, as this is the crucial year from which forecasts and harvest recommendations are made to managers. Estimated asymptotic approximations to the standard deviations of the biomass and movement parameters were also provided.

Boxplots of these key parameters were used for presentation, in which the box represents the inter-quartile range from the estimation models and the whiskers (short dashed vertical lines) represent the 95% confidence interval. Bias was considered significant if the value from the operating model was outside the 95% confidence intervals.

A.3 Results

Overall, the variability in the estimates of total and regional biomass in the last year was reduced when tagging datasets were integrated into the stock assessment model (Figure A.4). Estimated total and regional biomass for both the ASM and ITCAAN models were not biased (Figure A.4). The estimated standard deviations in total and regional biomass from ADMB were smaller for the tag recovery dataset uncertainty scenario (a) as compared to (b) for the ITCAAN models, but the estimated standard deviations from the ASM model were larger than either tag recovery dataset uncertainty scenario for the ITCAAN models (Figure A.5). Results were similar when tagging evenly between regions or proportionally to the BTS survey when considering the estimated standard deviations in total and regional biomass in the last year of the model (Figure A.5). Further, the estimated standard deviations in total and regional biomass were similar regardless of the length of the tag dataset time series (Figure A.5).

Parameter estimates for regional and seasonal movement were unbiased for both the ASM and ITCAAN models (Figure A.6). The estimated standard deviations in the movement parameter estimate for the NW region after the A season ($\theta_{\mathrm{NW}\rightarrow\mathrm{NW,\,A}}$) were similar for both the ASM and ITCAAN models (Figure A.7). Even though reduction in

the estimated standard deviations was small, for all other movement parameters the estimated standard deviations from ADMB were decreased in the ITCAAN models compared to the ASM model (Figure A.7). The largest differences in estimated standard deviations for movement parameters between the ASM and ITCAAN models occurred for tag recovery data uncertainty scenario (a) compared to scenario (b). In scenario (b) the estimated standard deviations in movement parameters from the ITCAAN models were only slightly lower than from the ASM model, but the general patterns were similar to results from scenario (a) (Figure A.7). When uncertainty was low in the tag return data in scenario (a) and tags were released evenly between regions, the estimated standard deviations of movement parameters were smaller when the time series of tag recovery data were longer (case II compared to case I). When tagging proportionally to the BTS survey the estimated standard deviations in movement parameters were similar regardless of the length of the tag recovery data (Figure A.7). Comparing between tag release methods, in tag recovery dataset uncertainty scenario (a) the estimated standard deviations of movement parameters were smaller when tagging evenly across regions as compared to tagging proportionally to the BTS survey; the same resulted for tag recovery data uncertainty scenario (b) but the differences were not as notable (Figure A.7).

A.4 Discussion

The quality of data, the quantity of data, and the soundness of assumptions all directly affect the accuracy of models and their predictions (NRC 1998). Both the ASM and ITCAAN models resulted in significantly unbiased estimates of parameters and

population biomass. This is a consequence of the operating and estimation models being the same, but it also shows there was apparently no appreciable parameter confounding even in the absence of tagging data. Our simulation study showed that an advantage of including tagging datasets into a spatially-explicit age-structured stock assessment model is primarily reduced variability in population biomass estimates and movement parameters. Even with a small-scale tagging study there was reduced variability in estimated total and regional biomass with the addition of tagging data. Thus, even if there is a large amount of uncertainty in observed data, the inclusion of tagging data will reduce the uncertainty in stock assessment estimates of biomass that are subsequently used for fishery management.

Variability was generally smaller in the movement parameter estimates when tagging data were included into the stock assessment model. When the movement parameter was near one (99%), however, variability was not affected, as shown with the movement parameter estimated after the A fishing season in the NW region (which is actually the probability of staying in the NW region). For all other movement parameters variability decreased when tagging data was used. We found that the number of years available for tag return data also influenced variability in the movement parameters. The differences in movement parameter variability when considering the length of tag return data were more evident when uncertainty was low in the tag recovery data. When uncertainty in the tag recovery data was larger, there was not a substantial improvement in movement parameter variability. Thus, the variability in the movement parameters decreased as the time series of tag data increased, but the magnitude of improvement in

movement parameter variability depended on the uncertainty in the tag recovery data and the tag release protocol.

Releasing tags evenly between regions provided the largest reduction in movement parameter variability. When tagging proportionally to the BTS survey catch we found that the variability in movement parameters was not largely affected by the number of years for which tag recovery data were available. However, variability in movement parameters was reduced when tagging evenly between regions as the number of years for which tagging datasets were available increased. Thus, we found that it is better to release tags evenly among regions rather than proportionally to abundance in this example. However, in this study we considered the simplified case in which complete mixing of tagged and non-tagged fish occurs immediately after the tags were released. Perhaps the spectrum of movement scenarios was too limited in this simulation; it may be that with alternative movement hypotheses such as random yearly movement that tagging proportionally may result in less uncertainty than tagging evenly. Also, the amount of mixing between tagged and untagged fish could identify differences between tag release protocols, a topic that requires further study. Our model is ideally suited to evaluate alternative experimental designs and protocols.

Pollock migration patterns could deviate in the future from historical patterns due to climate change forcing (e.g., Mueter and Litzow 2008). Temperature changes in the Bering Sea can affect pollock distribution between the U.S. and Russian Exclusive Economic Zones and the availability of this species to each fishery. Temperature changes can also affect the timing of pollock migration and the proportion of the stock within the

standard EBS survey area when the survey is conducted (Kotwicki et al. 2005). As the EBS pollock stock migrates further north as Bering Sea temperatures rise, these pollock may be missed in the EBS survey estimates, caught by the Russian fishery, and create management complications when determining the EBS total acceptable biological catch (Ianelli et al. 2001). In this simulation, movement was assumed constant across years. However, due to reduced uncertainty in total and regional biomass estimates when comparing the ASM and ITCAAN models, we hypothesize that tagging data would provide more accurate and precise estimates from stock assessment models when movement is stochastic, a hypothesis that will be tested in future research. If this is the case, then with tagging data management decisions would be made with less uncertainty, less likelihood of over-harvesting, and potentially higher sustained yields.

A.5 Acknowledgements

This work was supported by funding from the Cooperative Institute for Alaska Research (CIFAR)/ Alaska Fisheries Science Center (AFSC) and Alaska Sea Grant project number R/31-16. We thank Dr. Kalei Shotwell for constructing the map shown in Figure 1. We also thank referee Dr. André Punt and two anonymous referees for their helpful comments. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA.

References

Anganuzzi, A. 1996. Some technical issues regarding the study of fisheries interactions. *In* Status of Interactions of Pacific Tuna Fisheries in 1995, pp. 21-37. Ed. by R. Shomura, J. Majkowski, and R. Harman. FAO Fisheries Technical Paper no. 365.

Buckley, T. W., Greig, A., Ianelli, J., Livingston, P., and Walters, G. 2001. Spatial distribution and ontogenetic movement of pollock in the eastern Bering Sea. Poster presented to PICES, Alaska Fisheries Science Center, Seattle, WA.

Efron, B., and Tibshirani, R. J. 1993. An introduction to the Bootstrap. Chapman and Hall, New York.

Fournier, D. 1996. An introduction to AD Model Builder for use in nonlinear modeling and statistics. Nanaimo. Nanaimo, B.C., Otter Research Ltd., 56 pp.

Fu, C., and Quinn, T. J., II, 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Canadian Journal of Fisheries and Aquatic Sciences, 57(12): 2420–2432.

Haist, V. 1998. Integrated catch-age mark-recapture model: applications to B.C. sablefish stocks. *In* Fishery stock assessment models, pp. 469-494. Ed. by F. Funk, T. J.Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, and C. I. Zhang. Alaska Sea Grant Program Report AK-SG-98-01, University of Alaska, Fairbanks.

Hanselman, D., Lunsford, C. R., Fujioka, J. T., and Rodgveller, C. J. 2007. Alaska Sablefish Assessment for 2008. *In* Stock assessment and fishery evaluation report for 2008, pp. 195-312. North Pacific Fisheries Management Council. Anchorage, AK.

Heifetz, J., and Fujioka, J. 1991. Movement dynamics of tagged sablefish in the northeastern Pacific. Fisheries Research, 11: 355–374.

Hilborn, R., and Mangel, M. 1997. The Ecological Detective: Confronting Models with Data. Princeton, New Jersey, Princeton University Press.

Hilborn, R., and Walters, C. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. New York, NY, Chapman and Hall.

Ianelli, J. N., Barbeaux, S., Walters, G., Honkalehto, T, and Williamson, N. 2009. Eastern Bering Sea pollock stock assessment. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions, pp. 47-149. North Pacific Fisheries Management Council. Anchorage, AK.

Ianelli, J. N., Buckley, T., Honkalehto, T, Williamson, N., and Walters, G. 2001. Bering Sea-Aleutian Islands pollock assessment for 2002. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions, pp. 1-105. North Pacific Fisheries Management Council. Anchorage, AK.

Kotwicki, S., Buckley, T. W., Honkalehto, T., and Walters, G. 2005. Variation in the distribution of pollock (*Theragra chalcogramma*) with temperature and implications for seasonal migration. Fisheries Bulletin, 103: 574–587.

Maunder, M. 2001. Integrated Tagging and Catch-at-Age Analysis (ITCAAN): Model development and simulation testing. *In* Spatial Processes and Management of Marine Populations, pp. 123–146. Ed. by G. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and D. Witherell. Alaska Sea Grant College Program AK-SG-01-02, University of Alaska, Fairbanks.

McGarvy, R., Linnane, A. J., Feenstra, J. E., Punt, A. E., and Matthews, J. M. 2010. Integrating recapture-conditioned movement estimation into spatial stock assessment: A south Australian lobster fishery application. Fisheries Research, 105(2): 80–90.

Miller, S., Ianelli, J., and Quinn, T. J., II, 2008. Estimation of age-specific migration in an age-structured model. *In* Resiliency of Gadid Stocks to Fishing and Climate ChangeEd. by G. Kruse, K. Drinkwater, J. N. Ianelli, J. S. Link, D. L. Stram, V. Wespestad, and D. Woodby. Alaska Sea Grant College Program, AK-SG-08-01, University of Alaska, Fairbanks.

Miller, S. 2007. Estimating movement with a spatially-explicit stock assessment model of Eastern Bering Sea pollock (*Theragra chalcogramma*). M. S. thesis, School of Fisheries and Ocean Sciences. Fairbanks, AK, University of Alaska.

Mueter, F. J., and Litzow, M. A. 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecological Applications, 18(2): 309–320.

National Research Council (NRC) 1998. Improving Fish Stock Assessments. National Academy Press, Washington, DC.

Natural Resources Consultants, Inc., and Northwest Marine Technology, Inc. (N.R.C. & N.M.T.) 1996. Feasibility of tagging pollock, *Theragra chalcogramma*, with coded wire tags. N. M. F. S. Report prepared for Marine Resources Management Center and Resource Ecology and Fisheries Management Division, Seattle, WA.

Punt, A., Pribac, F., Walker, T. I., Taylor, B. L., and Prince, J. D. 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Marine Freshwater Research, 51(3): 205–220.

Quinn, T. J., II, and Deriso, R. 1999. Quantitative fish dynamics. New York, New York, Oxford University Press, Inc.

R Development Core Team 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Tuck, G., and Possingham, H. 1994. Optimal harvesting strategies for a metapopulation. Bulletin of Mathematical Biology, 56(1): 107–127.

Winter, A., Foy, R. J., and Trussell, M. 2007. Factors influencing the mortality of tagged pollock captures using a trawl net. North Pacific Research Board Project 506 Final Report (http://doc.nprb.org/web/05_prjs/506_final_report.pdf).

Table A.1. Years of spatially disaggregated data sources available for EBS pollock. Total yield included all years (1977 to 2005).

Region	Season	Bottom-trawl survey abundance	Echo-integrated trawl survey abundance	Fishery (Catch-at-age and Yield)
NW	A	1982-2004	1994, 1996, 1997, 1999, 2000, 2002	1977-2005 (Missing 1988-1990, 1997)
	B			1977-2005 (Missing 1988-1990)
SE	A	1982-2004	1994, 1996, 1997, 1999, 2000, 2002	1977-2005 (Missing 1988-1990)
	B			1977-2005 (Missing 1988-1990)

Table A.2. Input values for σ used in the log-normal distribution. The letters 'na' under region or season stand for 'not applicable' under that column.

Data sources	Region	Season	σ
Fishery yield	NW	A	0.05
		B	0.05
	SE	A	0.05
		B	0.05
Total Fishery Yield	na	na	0.025
Fishery catch-at-age	NW	A	0.3
		B	0.2
	SE	A	0.2
		B	0.2
Bottom trawl survey	NW	na	0.2
	SE	na	0.2
Echo-integrated trawl survey	NW	na	0.2
	SE	na	0.2

Table A.3. Standard deviations considered in the two uncertainty scenarios for tag recovery data.

Data source	Region	Season	a	b
Tag Recoveries	NW	A	0.05	0.7
		B	0.05	0.2
	SE	A	0.05	0.1
		B	0.05	0.1

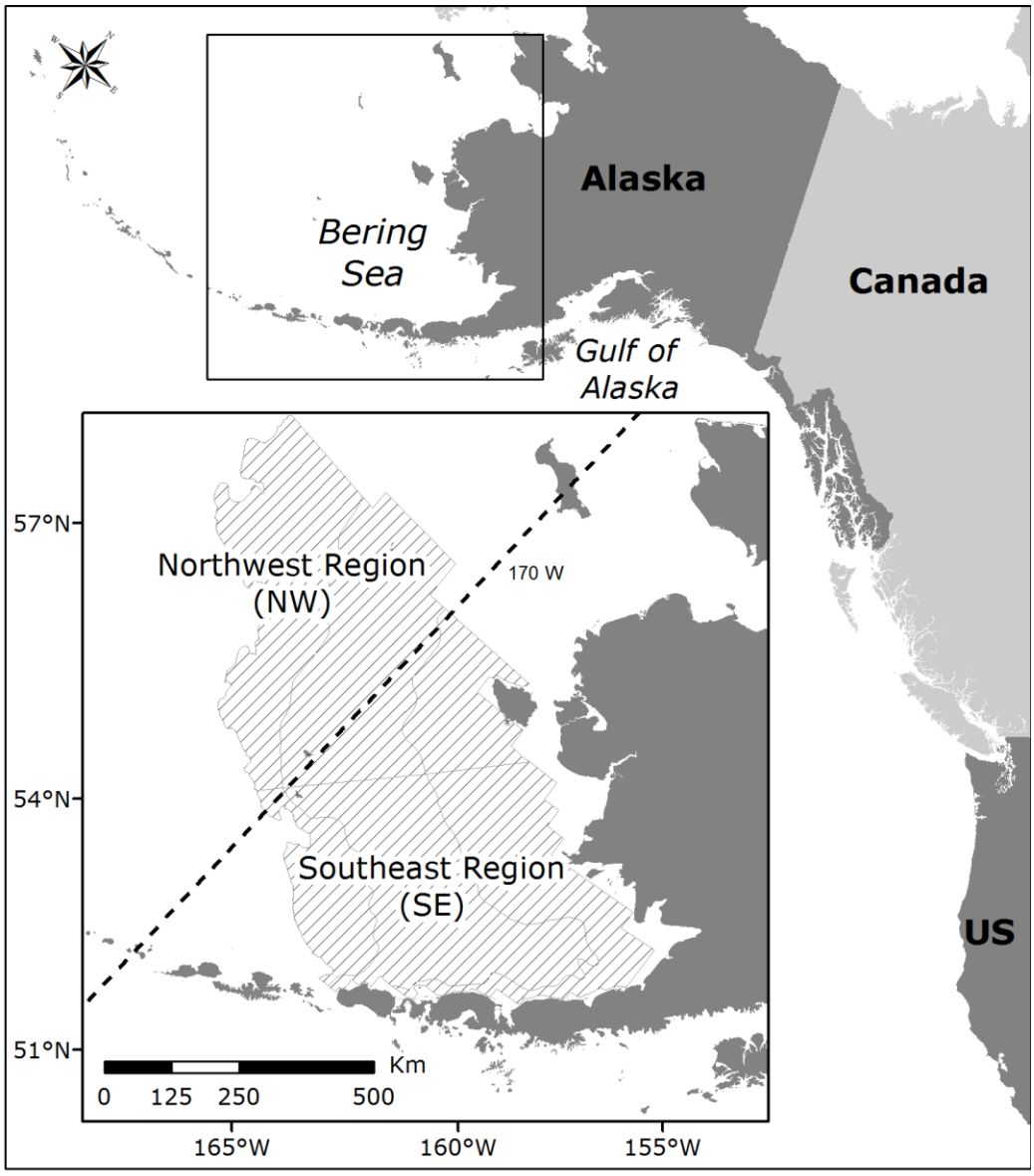

Figure A.1. Map of the eastern Bering Sea. Illustration showing the Northwest and Southeast regions of the eastern Bering Sea considered in the spatial modeling of walleye pollock.

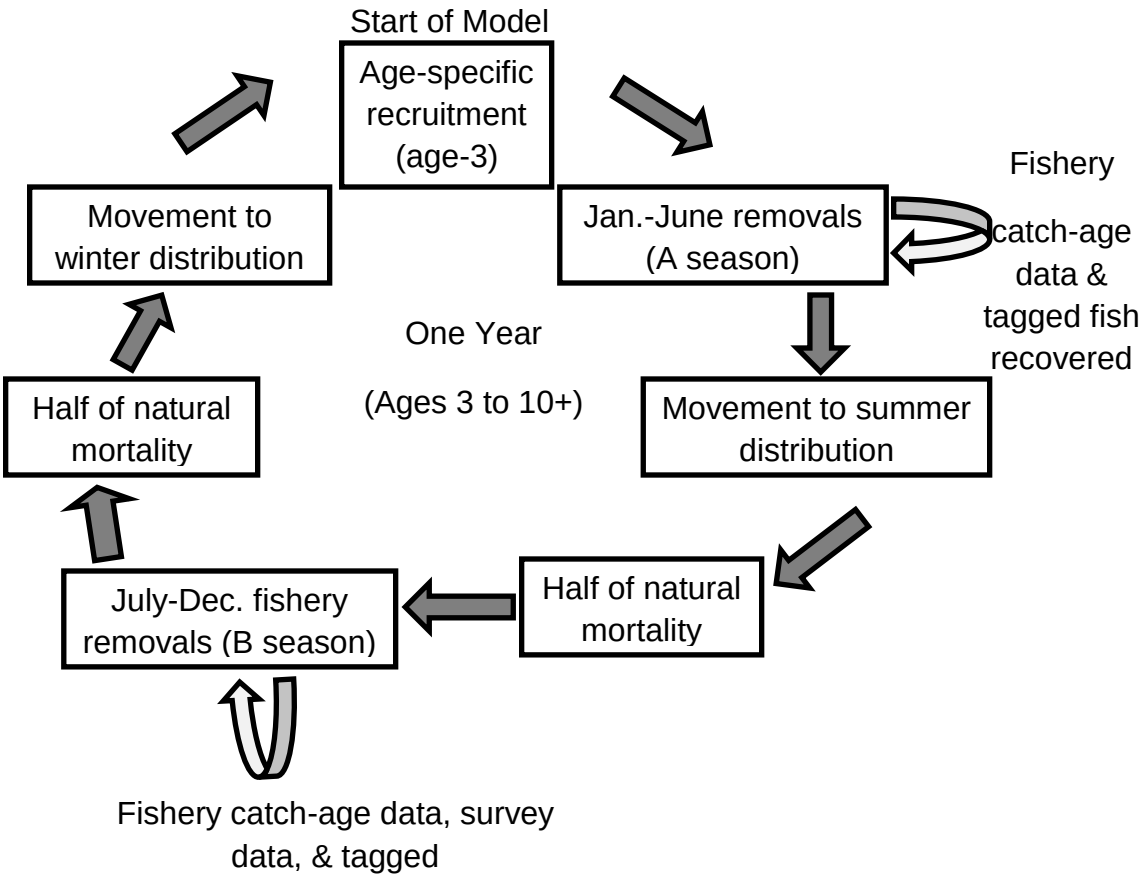

Figure A.2. Depiction of the two-area, two-season model. Applied to pollock in the eastern Bering Sea, showing the processes affecting the population. Figure modified from Miller et al. 2008.

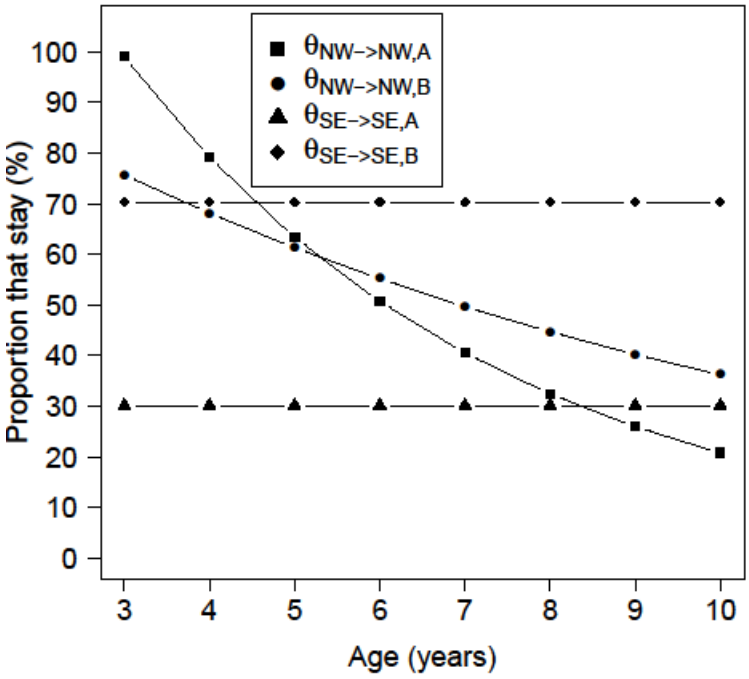

Figure A.3. Movement parameter estimates, by age. Modified from Miller et al. (2008).

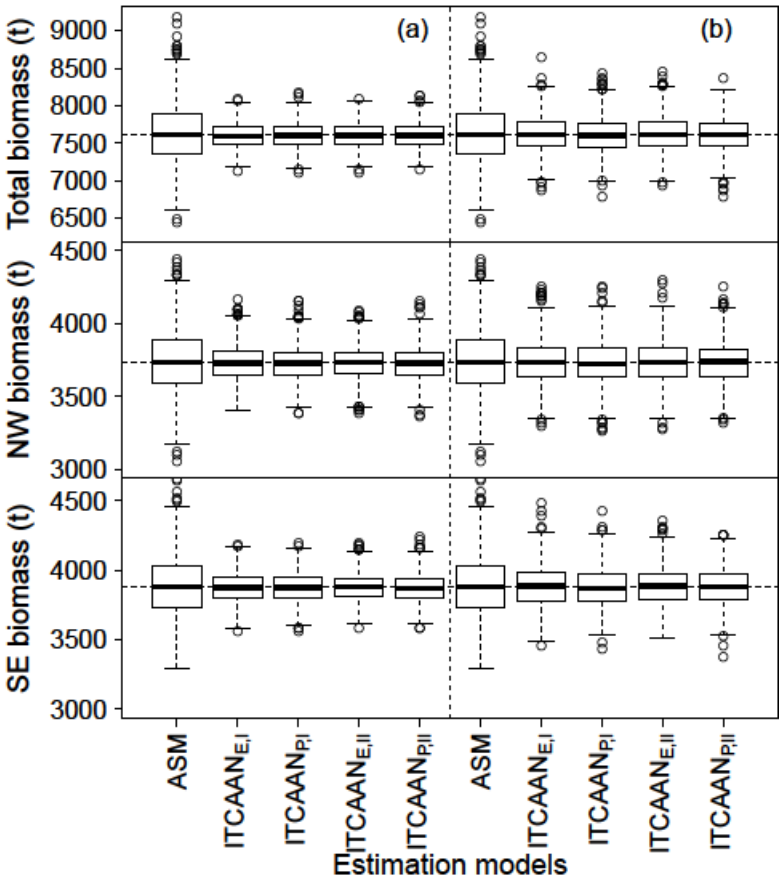

Figure A.4. Total, NW, and SE biomass estimated by the ASM and ITCAAN models. Tag recovery data uncertainty scenario (a) is shown in the left panel, (b) in the right. The box represents the inter-quartile range from the estimation models, the whiskers (short dashed vertical lines) represent the 95% confidence interval, and values outside of the 95% confidence intervals are shown with open circles.

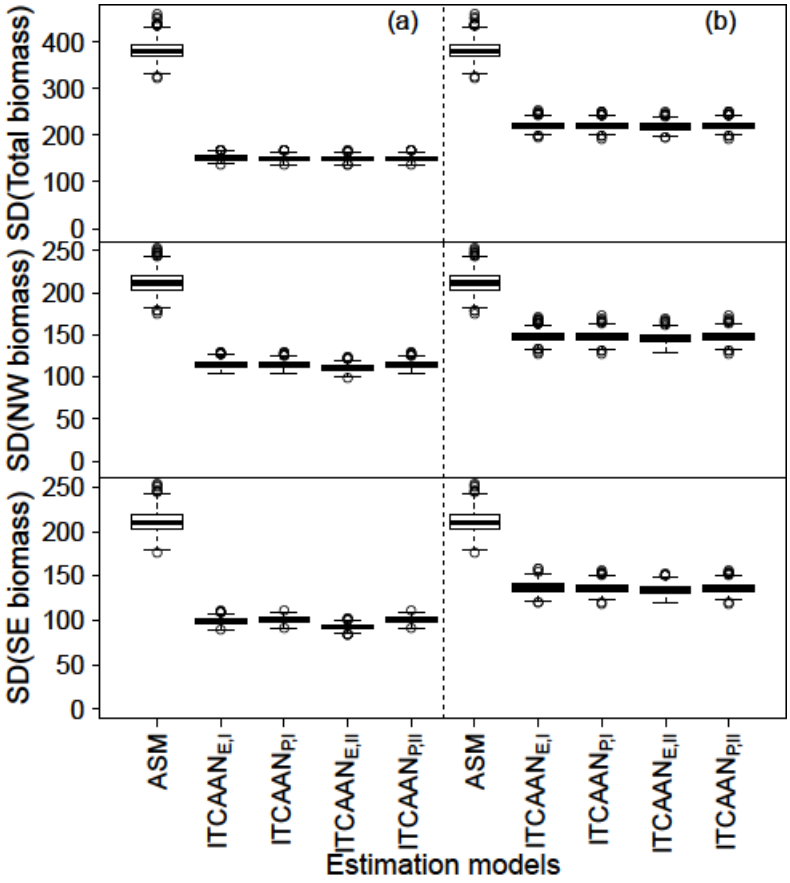

Figure A.5. Standard deviations (SD) in total, NW, and SE biomass. Estimated with ADMB for the ASM and ITCAAN models. Tag recovery data uncertainty scenario (a) is shown in the left panel, (b) in the right. The box represents the inter-quartile range from the estimation models, the whiskers (short dashed vertical lines) represent the 95% confidence interval, and values outside of the 95% confidence intervals are shown with open circles.

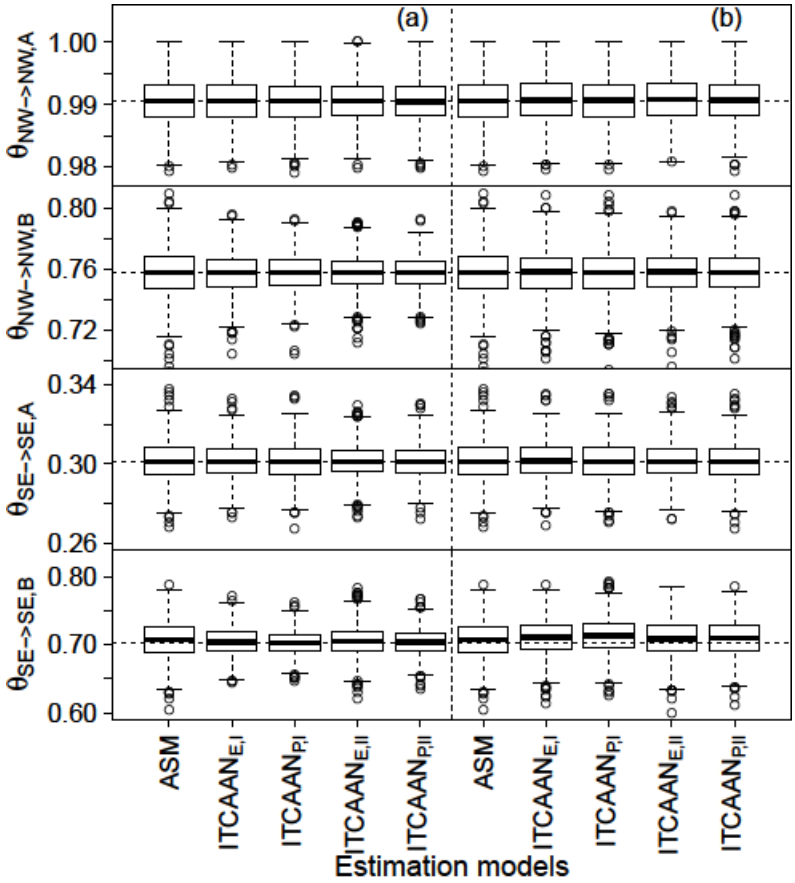

Figure A.6. Estimated movement parameters from the ASM and ITCAAN models. Tag recovery data uncertainty scenario (a) is shown in the left panel, (b) in the right. The box represents the inter-quartile range from the estimation models, the whiskers (short dashed vertical lines) represent the 95% confidence interval, and values outside of the 95% confidence intervals are shown with open circles.

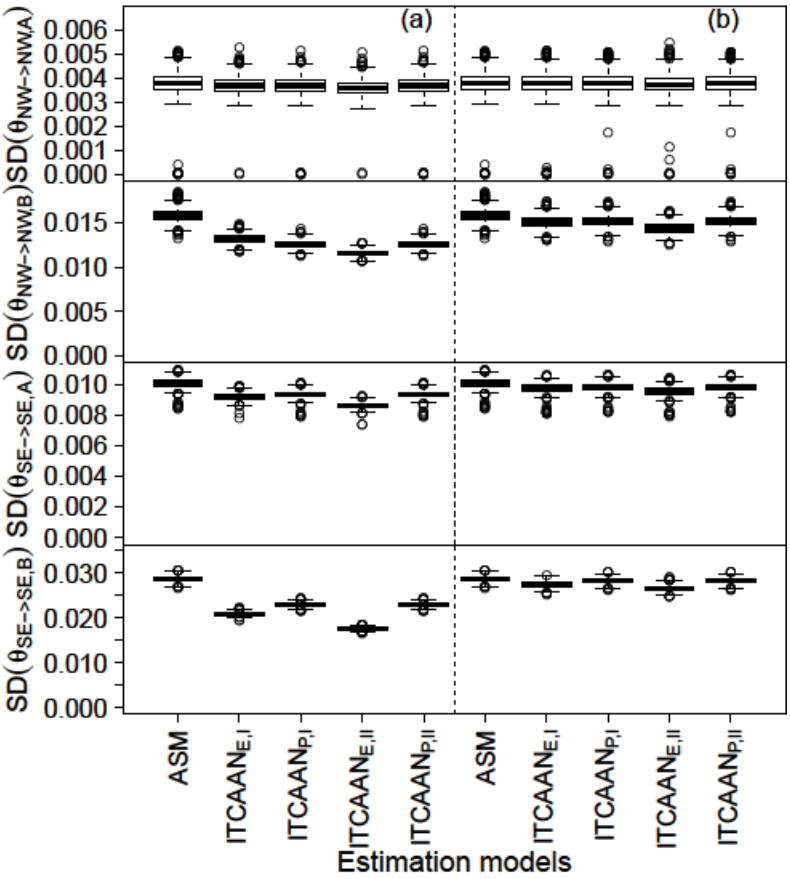

Figure A.7. Standard deviations (SD) in movement parameters. Estimated by ADMB for the ASM and ITCAAN models. Tag recovery data uncertainty scenario (a) is shown in the left panel, (b) in the right. The box represents the inter-quartile range from the estimation models, the whiskers (short dashed vertical lines) represent the 95% confidence interval, and values outside of the 95% confidence intervals are shown with open circles.